

00 per volume
le numbers \$2.00

QUARTERLY

February, 1931
Volume II, Number 1

THE JOURNAL OF SOCIAL PSYCHOLOGY

Political, Racial,
and Differential Psychology

EDITED BY
JOHN DEWEY
CARL MURCHISON

Floyd H. Allport
Syracuse University

Franz Boas
Columbia University

Trigant Burrow
New York City

Havelock Ellis
London

Franklin H. Giddings
Columbia University

G. Dawes Hicks
University of London

A. Hrdlička
Smithsonian Institution

Charles H. Judd
University of Chicago

David Katz
Universität Rostock

K. Kornilov
Gosudarstvennii Institut
Experimentalnoi
Psikhologii, Moskva

L. Lévy Bruhl
Université de Paris

Mark A. May
Yale University

William McDougall
Duke University

Raymond Pearl
The Johns Hopkins University

Walter B. Pillsbury
University of Michigan

A. T. Poffenberger
Columbia University

Bertrand Russell
Petersfield, England

Edward Sapir
University of Chicago

Carl E. Seashore
University of Iowa

Giuseppe Sergi
R. Università di Roma

L. L. Thurstone
University of Chicago

Clark Wissler
Yale University

Kimball Young
University of Wisconsin

Luberta M. Harden, Ph.D.
Assistant Editor

Worcester, Massachusetts
Copyright, 1931, by Clark University

Entered as second-class matter January 10, 1930, at the post-office at
Worcester, Mass., under Act of March 3, 1879

Journal of General Psychology

EDITED BY

CARL MURCHISON, *Clark University*

Frank Angell, *Stanford University*; F. C. Bartlett, *Cambridge University*; V. M. Borovski, *Gosudarstvennii Institut Eksperimentalnoi Psikhologii, Moskva*; G. S. Brett, *University of Toronto*; Karl Bühler, *Universität, Wien*; Harvey Carr, *University of Chicago*; W. J. Crozier, *Harvard University*; Raymond Dodge, *Yale University*; James Drever, *University of Edinburgh*; Georges Dumas, *Université de Paris*; Knight Dunlap, *The Johns Hopkins University*; Shepherd Ivory Franz, *University of California at Los Angeles*; H. L. Hollingworth, *Columbia University*; T. A. Hunter, *Victoria University College, New Zealand*; Pierre Janet, *Université de Paris*; F. Kiesow, *Regia Università degli Studi, Torino*; Otto Klemm, *Universität, Leipzig*; Felix Krueger, *Universität, Leipzig*; Matataro Matsumoto, *Tokyo Imperial University*; William McDougall, *Duke University*; A. Michotte, *Université de Louvain*; Walter Miles, *Stanford University*; John Paul Nafe, *Clark University*; Kanae Sakuma, *Kiushiu Imperial University, Fukuoka*; A. L. Schniermann, *Bekhterev Refleksologicheskii Institut po Ixucheniya Mozga, Leningrad*; C. E. Spearman, *University of London*; Edward C. Tolman, *University of California*; John B. Watson, *New York City*; Albert P. Weiss, *Ohio State University*; H. P. Weld, *Cornell University*; Raymond H. Wheeler, *University of Kansas*; R. W. Wilcocks, *Universiteit van Stellenbosch, Suidafrika*; Clark Wissler, *Yale University*; Robert M. Yerkes, *Yale University*.

Quarterly, published in January, April, July, and October. Devoted primarily to experimental, theoretical, clinical, and historical psychology. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to the general editorial office. All subscriptions and business communications should be sent directly to the Clark University Press. Five to six hundred pages annually. Per annum, \$7.00; single numbers, \$2.00.

The Pedagogical Seminary and Journal of Genetic Psychology

EDITED BY

CARL MURCHISON, *Clark University*

John E. Anderson, *University of Minnesota*; Charlotte Bühler, *Universität, Wien*; William H. Burnham, *Clark University*; Cyril Burt, *University of London*; Ed. Claparède, *Université de Genève*; Edmund S. Conklin, *University of Oregon*; Sante De Sanctis, *R. Università di Roma*; Arnold Gesell, *Yale University*; William Healy, *Judge Baker Foundation, Boston*; Leta S. Hollingworth, *Teachers College, Columbia University*; Walter S. Hunter, *Clark University*; Buford Johnson, *The Johns Hopkins University*; Harold E. Jones, *University of California*; Truman L. Kelley, *Stanford University*; Yoshihide Kubo, *Hiroshima Normal College*; K. S. Lashley, *University of Chicago*; A. R. Luria, *Akademiya Kommunisticheskogo Vospitaniya im. N. K. Krupskoi, Moskva*; Toshio Nogami, *Kyoto Imperial University*; Ivan P. Pavlov, *Gosudarstvennii Institut Eksperimentalnoi Meditsini, Leningrad*; Henri Piéron, *Université de Paris*; William Stern, *Hamburgische Universität*; Calvin P. Stone, *Stanford University*; Lewis M. Terman, *Stanford University*; Godfrey Thomson, *University of Edinburgh*; E. L. Thorndike, *Teachers College, Columbia University*; C. J. Warden, *Columbia University*; John B. Watson, *New York City*; Helen Thompson Woolley, *Teachers College, Columbia University*.

Quarterly, published in March, June, September, and December. Devoted to child behavior, animal behavior, and comparative psychology. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to the general editorial office. All subscriptions and business communications should be sent directly to the Clark University Press. Six hundred pages annually. Per annum, \$7.00; single numbers, \$2.00. Complete sets from 1891 at \$7.00 and \$9.00 per volume, plus transportation.

CLARK UNIVERSITY PRESS

Worcester, Massachusetts

U. S. A.

THE JOURNAL OF SOCIAL PSYCHOLOGY

Political, Racial,
and Differential Psychology

EDITED BY

JOHN DEWEY
CARL MURCHISON

FLOYD H. ALLPORT
Syracuse University

FRANZ BOAS
Columbia University

TRIGANT BURROW
New York City

HAVELOCK ELLIS
London

FRANKLIN H. GIDDINGS
Columbia University

G. DAWES HICKS
University of London

A. HRDLÍČKA
Smithsonian Institution

CHARLES H. JUDD
University of Chicago

DAVID KATZ
Universität Rostock

K. KORNILOV
Gosudarstvennii Institut
Experimentalnoi
Psikhologii, Moskva

L. LÉVY BRUHL
Université de Paris

MARK A. MAY
Yale University

WILLIAM McDUGALL
Duke University

RICHARD MÜLLER-FREIBENFELS
Pädagog. Akademie, Stettin

RAYMOND PEARL
The Johns Hopkins University

WALTER B. PILLSBURY
University of Michigan

A. T. POFFENBERGER
Columbia University

BERTRAND RUSSELL
Petersfield, England

EDWARD SAPIR
University of Chicago

CARL E. SEASHORE
University of Iowa

GIUSEPPE SERGI
R. Università di Roma

L. L. THURSTONE
University of Chicago

CLARK WISSLER
Yale University

KIMBALL YOUNG
University of Wisconsin

LUBERTA M. HARDEN, PH.D.
Assistant Editor

VOLUME II

1931

Published quarterly by the
CLARK UNIVERSITY PRESS
Worcester, Massachusetts
U. S. A.

Copyright, 1931,
by
CLARK UNIVERSITY

TABLE OF CONTENTS

ARTICLES

CECELIA G. ALDRICH AND EDGAR A. DOLL	
Problem-Solving among Idiots: The Use of Implements	306-336
HERBERT S. CONRAD	
The Measurement of Adult Intelligence, and the Requisites for a General Intelligence Test	72-86
HERBERT S. CONRAD AND HAROLD E. JONES	
Psychological Studies of Motion Pictures: V. Adolescent and Adult Sex Differences in Immediate and Delayed Re- call	433-459
EDGAR A. DOLL AND CECELIA G. ALDRICH	
Problem-Solving among Idiots: The Use of Implements	306-336
HANNA F. FATERSON	
Organic Inferiority and the Inferiority Attitude	87-101
K. C. GARRISON AND MARGARET MANN	
A Study of the Opinions of College Students	168-178
J. P. GUILFORD	
Racial Preferences of a Thousand American University Stu- dents	179-204
ELIZABETH B. HURLOCK	
The Psychology of Incentives	261-290
HERBERT H. JASPER	
Is Perseveration a Functional Unit Participating in All Be- havior Processes?	28-51
HAROLD E. JONES AND HERBERT S. CONRAD	
Psychological Studies of Motion Pictures: V. Adolescent and Adult Sex Differences in Immediate and Delayed Re- call	433-459
C. M. LOUITT	
Racial Comparisons of Ability in Immediate Recall of Lo- gical and Nonsense Material	205-215
MARGARET MANN AND K. C. GARRISON	
A Study of the Opinions of College Students	168-178
JULIUS B. MALLER	
Size of Family and Personality of Offspring	3-27
H. MELTZER	
The Forgetting of Pleasant and Unpleasant Experiences in Relation to Intelligence and Achievement	216-229
ELINOR NOETZEL AND IRA S. WILE	
A Study of Birth Order and Behavior	52-71
RAYMOND PEARL	
Some Notes on the Census of Religious Bodies, 1926	417-432
ROBERT SAUDEK	
Writing Movements as Indications of the Writer's Social Behavior	337-373
KEITH SWARD	
Temperament and Religious Experience	374-396

L. L. THURSTONE	
The Indifference Function	139-167
L. L. THURSTONE	
Influence of Motion Pictures on Children's Attitudes . . .	291-305
RALPH K. WHITE	
The Versatility of Genius	460-489
IRA S. WILE AND ELINOR NOETZEL	
A Study of Birth Order and Behavior	52-71

SHORT ARTICLES AND NOTES

ALICE BEHNER AND PAUL R. FARNSWORTH	
A Note on the Attitude of Social Conformity	126-128
WILLIAM W. BIDDLE	
The Relationship Between Knowledge and a Measure of Autistic Thinking on Certain International Problems . . .	493-496
PAUL L. BOYNTON	
Are Girls Superior to Boys in Visual Memory?	496-500
MASON N. CROOK AND LUBERTA HARDEN	
A Quantitative Investigation of Early Memories	252-255
D. D. DROBA	
Czechoslovak Psychology of Testing	238-245
PAUL R. FARNSWORTH AND ALICE BEHNER	
A Note on the Attitude of Social Conformity	126-128
THOMAS R. GARTH, KUNIHEI IKEDA, AND ROY M. LANGDON	
The Color Preferences of Japanese Children	397-402
LUBERTA HARDEN AND MASON N. CROOK	
A Quantitative Investigation of Early Memories	252-255
KUNIHEI IKEDA, THOMAS R. GARTH, AND ROY M. LANGDON	
The Color Preferences of Japanese Children	397-402
ROY M. LANGDON, THOMAS R. GARTH, AND KUNIHEI IKEDA	
The Color Preferences of Japanese Children	397-402
C. W. LUH AND T. M. WU	
A Comparative Study of the Intelligence of Chinese Children on the Pintner Performance and the Binet Tests	402-408
EDWARD J. V. K. MENGE	
Psychological Reactions of Biologists: A Cross-Section of Present-Day Biological Thought	102-114
JOHN T. METCALF	
Empathy and the Actor's Emotion	235-238
HERBERT MOORE	
A Comparison of Linguistic and Non-Linguistic Ability . .	245-252
FRANCIS F. POWERS	
The Influence of Intelligence and Personality Traits upon False Beliefs	490-493
STUART M. STOKES AND ELMER D. WEST	
Sex Differences in Conversational Interests	120-126
L. L. THURSTONE	
The Measurement of Change in Social Attitude	230-235
JOHN N. WASHBURN	
Definitions in Character Measurement	114-119
ELMER D. WEST AND STUART M. STOKES	
Sex Differences in Conversational Interests	120-126

T. M. WU AND C. W. LUH

A Comparative Study of the Intelligence of Chinese Children on the Pintner Performance and the Binet Tests . . .	402-408
First International Congress of Religious Psychology . . .	129

BOOKS

J. W. BRIDGES

Psychology, Normal and Abnormal (P. H. Ewert) . . .	409-413
---	---------

HAROLD D. LASSWELL

Psychopathology and Politics (Raymond R. Willoughby) . . .	256-257
--	---------

MARGARET MEAD

Growing Up in New Guinea (Raymond R. Willoughby) . . .	130-136
--	---------

MARGARET MEAD

Coming of Age in Samoa (Raymond R. Willoughby) . . .	130-136
--	---------

CARL MURCHISON [Ed.]

A Handbook of Child Psychology (Frank S. Freeman) . . .	501-502
---	---------

SIZE OF FAMILY AND PERSONALITY OF OFFSPRING*¹

*From the Institute of School Experimentation, Teachers College,
Columbia University*

J. B. MALLER

Do children of large families differ in intelligence and character from those of small families? Does the "only" child have its own characteristics?

The purpose of this research was to study the exact relationship between the number of children in the family and such intellectual factors as brightness, moral knowledge, and cultural background; such measurable factors of character as cooperativeness, self-control, and honesty; ratings on personality by classmates and teachers; intelligence of parents; and socio-economic environment.

SOURCE OF DATA

The Character Education Inquiry, sponsored by the Institute of Social and Religious Research, and conducted at Teachers College, Columbia University, gathered extensive information concerning the children of three schools, referred to as *E*, *W*, and *L*. For each child of these three populations there were available a vast number of scores on factors of intelligence and character. The size of the family, i.e., the number and names of a child's brothers and sisters, was among the items of information for 802 children.

PREVIOUS STUDIES

The relationship between children's intelligence and the number of children in their families has been recorded in a number of studies. Chapman and Wiggins (1) found a correlation of $-.33 \pm .024$

*Accepted for publication by A. T. Poffenberger of the Editorial Board and received in the Editorial Office April 28, 1930.

¹The writer is indebted to Dr. Hugh Hartshorne and Dr. Mark May of Yale University, Directors of the Character Education Inquiry, for permission to use their data for this study. Dr. Otis W. Caldwell and Dr. Albert Poffenberger of Columbia University have offered many valuable suggestions.

between size of family and IQ on the N.I.T. The scores on the arithmetic section of the Stanford Achievement Test were also correlated with the size of family. The correlation was $-.26 \pm .03$, slightly lower than the correlation with score on the N.I.T.

They found an interesting relationship between the parents' education and family size. Parents, both of whom had a high school education, had an average of 3.3 ± 2.3 children per family, while parents, neither of whom attended high school, had an average of 4.90 ± 2.8 per family. Where only the father or only the mother had a high school education, the average was 4.20 ± 2.9 , and 3.40 ± 2.6 , respectively.

Sutherland and Thomson (14) report a correlation of $-.20$ between the two variables. Slawson (12) found a negative correlation between intelligence and size of family for children in institutions for delinquents ($-.11$ for children in a cosmopolitan institution, and $-.23$ for children in a Jewish institution for delinquents).

Pearson and Moul (10) report that the correlation ratio they found between intelligence and size of family was not higher than the ratio obtainable by pure chance. Their study, however, was based on children of extremely large families and a highly selected group.

Maller (8) reported a significant correlation between children's cooperativeness and the number of brothers and sisters.

The present research deals with the relation between family size and various measures of personality.

THE POPULATION STUDIED

The children for whom data were obtained came from three different schools. School E is located in a residential neighborhood in New Haven. The social and economic level is above average and the parents are of professional and business status. School W is the only public school of Walden, a town of 5000 located about 65 miles from New York. All the children of the town of Grades V-VIII were tested. The parents, 80% of whom are native-born, are mainly of English extraction. It is a typical small community with a population clustering about the average in social and economic status. School L is a public school in a poor, under-privileged district in New Haven. The parents are mostly unskilled laborers, two-thirds of them foreign-born. Of the latter 41% were born in Italy, 28% in Slavic countries, and 15% in Ireland.

The total population consisted of 802 children, composed of the lower-, middle-, and above-average social levels in about equal proportions. The ages of the children ranged from 9 to 17, with an average of 12.3 years. The "young families" that usually complicate the relationship with size of family were thus eliminated to some extent.

RESULTS

The Size of Families in the Population Studied. Our population was composed of three different social levels. What is the average size of family in each of these groups? This is shown in Table 1, which gives the distribution of the size of family for each of our three school populations.

TABLE 1
DISTRIBUTION OF SIZE OF FAMILY FOR POPULATIONS E, W, L

No. of children per family	E	School W	L	Total
1	50	54	42	146
2	52	72	32	156
3	59	66	47	172
4	26	44	42	112
5	20	12	37	69
6	7	19	33	59
7	2	7	31	40
8	1	3	9	13
9	2	2	8	12
10	1	0	3	5
11		3	1	4
12			1	1
Mean	2.85	3.08	4.26	3.44
S.D.	1.64	1.88	2.38	2.11
<i>V</i>	57.50	61.00	55.80	61.30
<i>N</i>	220	282	287	789

In School E there were 50 children who came from families of one child, 52 children came from families of two children, etc. The average number of children per family, the standard deviations, the coefficients of variation, and the number of cases are also given.

It appears from Table 1 that in our total population there were 3.44 children per family. There were considerable differences between the three populations. The children of School E, superior social level, came from homes of 2.85 children per family; those of School W, average social level, from homes of 3.08 children per

family; and those of School L came from homes of 4.26 children per family. These averages are comparatively low.

Slawson reports for New York City an average of 4.49, with an *S.D.* of 2.14; for the cities over 25,000 an average of 4.69 and an *S.D.* of 2.34; and for cities under 25,000 an average of 4.73 and an *S.D.* of 2.40. Those averages are based on populations somewhat differently composed from ours. In our population the lower social group constituted only one-third of the total, while in the general population they probably constitute a much larger proportion.

Social Level and Size of Family. The difference in number of children per family (Table 1) between Schools W and L was 1.18, the *S.D.* of the difference was .178, and the ratio of the difference to its *S.D.* was 6.62. The difference between Schools W and E was 0.23, the *S.D.* was .1556, and the ratio of the difference to its *S.D.* was 1.48; while the difference between Schools E and L was 1.41, its *S.D.* was .178, and the ratio of the difference to its *S.D.* was 7.91. Thus, we see that the size of the family in School L (low social level) was significantly larger than in either of the other two schools. There was a difference in size of family between Schools E (upper middle class) and W (average), but the difference was statistically insignificant.

Intelligence and Size of Family. A Thorndike Intelligence Test was administered to the total population. The test is composed of four parts: information, vocabulary, arithmetic, and sentence completion. The obtained intelligence score is similar to, though not identical with, the intelligence quotient. It indicates a child's deviation from the mean of his age group in terms of the *S.D.* of the distribution.

In view of the significance of the problem, we present in Chart 1 the scattergram showing the intelligence scores by size of family.

The intelligence scores of these 789 children ranged from 23 to 75 with an average score of 47.76 and an *S.D.* of 8.74, and the number of children per family ranged from 1 to 10 (there were a few families having more than 10 but they were all added to the 10-group), with a mean of 3.44 and an *S.D.* of 2.11. The correlation between intelligence and the number of children in the family is $-.2076 \pm .02$. It indicates a slight but statistically significant tendency for intelligence to decrease with size of family.

The relationship between the two variables becomes clearer if we divide the total population into 10 groups, according to the number

CHART 1
DISTRIBUTION OF INTELLIGENCE BY SIZE OF FAMILY

	N	146	156	172	112	69	59	40	13	12	10	789
Intelligence Score	75	1		1								2
	73			1								1
	71		1	2	1							4
	69	2	1	1								4
	67	1	2	6		1	1				1	12
	65	3	1	3	1	1						9
	63	3	10	1	4		1					19
	61	4	4	4	3	1						16
	59	9	6	7	4	3		1				30
	57	9	8	10	6	2	2	1		1		39
	55	8	11	9	10	4	1	1				44
	53	5	10	7	3	4	2	1	1		1	34
	51	23	26	21	10	5	8	3	2			98
	49	9	9	15	12	9	5	2	1	1	1	64
	47	9	18	17	8	7	6	2	2	1		70
	45	15	10	15	7	5	8	4	1	2	2	69
	43	15	8	15	12	8	10	3		2	1	74
	41	8	12	11	9	4	4	6	3	1		58
	39	7	8	8	5	4	3	5	1	1	1	43
	37	3	5	5	6	2	2	3	2	2	1	31
	35	3	4	6	2	4	2	4				25
	33	7	1	4	5	2	2	2		1	1	25
	31	1	1	1	2	1	2	1				9
	29											0
	27			1	1	1					1	4
	25	1			1	1		1				4
	23			1								1
		1	2	3	4	5	6	7	8	9	10	
		Number of children in family										

TABLE 2
AVERAGE INTELLIGENCE BY SIZE OF FAMILY

Number of children per family	1	2	3	4	5	6	7	8	9	10
Average intel- ligence	48.25	49.73	48.88	47.30	46.28	45.47	42.30	44.54	43.00	43.80
Number of cases	146	156	162	112	69	59	40	13	12	10

of children in the families whence they come, and then find the average intelligence of each group. Table 2 shows the means of intelligence for children grouped on the basis of the size of family.

Three important facts are revealed in the above table.

1. Those children who have only one brother or sister, who come from families of two children, were of highest intelligence.

2. Children coming from successively larger families scored lower in intelligence with considerable consistency.

3. "Only" children, coming from families of one child, scored above average but lower than those who come from families of two or three children.

Curvilinearity of Relationship. The above correlation was the Pearson product-moment coefficient. In view of the fact that the regression line was not linear, for the means first increased and then decreased, we computed the correlation ratio, *eta*, between the two variables. The *eta* of intelligence on size of family was $.23 \pm .02$.

Consistency of Relationship. The correlation was consistently negative when computed separately for the different schools. In School L, lowest in average intelligence, the correlation was $-.1978 \pm .04$. In School W the correlation was negative but within its probable error. The highest negative relationship between intelligence and family size was found in the poorest population, L, and the lowest relationship was found in the "average" population, W. In population L there is evidently a greater tendency for the large family to be associated with general inferior conditions, while in the well-established community of W there are probably other unisolated factors involved.

Moral Knowledge of Children and Size of Family. The Character Education Inquiry devised and administered a number of moral knowledge tests. The scores of one of those tests were used in this study, the Good Citizenship Test. It is composed of 50 questions testing the child's knowledge of what is the most sensible, helpful, and useful thing to do in situations involving social adjustment.

The correlation between score on the Good Citizenship Test and size of family was computed for the children of Schools E and L. They were $-.1601 \pm .04$ and $-.1289 \pm .04$, respectively. For the two schools combined, the correlation rose to $-.2346 \pm .03$. The relationship between the two variables is definitely negative. Table 3 shows the average moral knowledge score of each group classified according to the number of children in the family.

TABLE 3
SCORE ON GOOD CITIZENSHIP BY SIZE OF FAMILY

Number of children in family	1	2 or 3	4 or 5	6 or 7	8 or 9	10 or more
Score on Good Citizenship Test	59.6	58.4	53.5	50.1	45.2	45.0
Number of cases	85	199	128	74	20	8

This table reveals that the average moral knowledge score decreased consistently with the number of children in the family. The "only"-child group scored highest, followed closely by those who come from homes of 2 or 3 children. The scores of the succeeding groups decreased continuously without exception. The relationship between moral knowledge and size of family was more pronouncedly negative. The relationship was more linear than in the case of intelligence. There the "only"-child group, though above the average, scored lower than those of the 2-or-3-child families. It is likely that the "only" child receives more of the parent's attention and has more opportunity to learn the most adequate responses to the social situations presented in the Good Citizenship Test and therefore scored highest on the test involving such responses.

Cultural Background and Size of Family. A measure of a child's cultural background was obtained by means of the Burdick Apperception Test. Three elements enter in a child's score: (a) Economic factors, such as the parent's occupation, income, the child's recreational facilities, etc. (b) Cultural factors, such as familiarity with music, art, literature, knowledge of etiquette, manner of spending leisure time, etc. (c) Ethical factors, such as parental supervision, character of adult members of family, attitude of members of the family toward one another.

Table 4 gives the average scores on this measure of home background by the number of children in the family. These figures reveal that the highest score on cultural background was made by children coming from families of 2 or 3 children, followed by the group of "only" children. Those coming from homes of 6 or more scored somewhat higher than those who come from families of 4 or 5. Children of small families (3 or less) scored higher than those of

large families (4 or more). The correlation between size of family and cultural background was $-.112 \pm .04$, indicating a low but significantly negative relationship.

TABLE 4
CULTURAL BACKGROUND SCORE BY SIZE OF FAMILY
(School W only)

Number of children in family	1	2 or 3	4 or 5	6 or more
Score	142.8	149.6	133.5	138.0
N	49	126	48	33

Cooperativeness and Size of Family. Is there any relationship between a child's readiness to cooperate, to be of service, and the number of children in his family? The children of Schools E, W, and L were given five different tests measuring a child's readiness to be helpful and cooperative. The relationship to size of family will be presented for each of those tests.

1. *The Free Choice Test.* The score on this test indicates whether or not a child had chosen, of his own accord, to work on a group project on which no individual names were written, rather than to work for a personal prize and recognition. Table 5 gives the percentage of children cooperating on this test by size of family. This table reveals that in each one of the schools, the group of "only" children had the smallest percentage of cooperative children. Those who come from families of 4 or 5 children (6 or 7 in School W) had the highest percentage of cooperative children.

TABLE 5
FREE CHOICE COOPERATION AND SIZE OF FAMILY

Number of children in family	School E		School W		School L	
	N	Percentage cooperating	N	Percentage cooperating	N	Percentage cooperating
1	53	61	54	63	46	71
2 or 3	109	69	139	69	80	79
4 or 5	28	75	41	83	79	83
6 or 7	18	66	32	84	65	72
8 or more	13	62	15	80	25	76

The Free Choice Test yielded also a score of the extent of cooperation. There were seven opportunities of choosing to cooperate, and the scores ranged from 0 to 7.

TABLE 6
SCORE ON FREE CHOICE BY SIZE OF FAMILY

Number of children in family	1	2 or 3	4 or 5	6 or 7	8 or 9	10 or more
Average score on Free Choice	2.2	2.8	3.0	2.3	2.2	1.7
N	153	323	178	100	24	14

TABLE 7
EFFICIENCY COOPERATION SCORES BY SIZE OF FAMILY

Number of children	1	2	3	4	5	6	7	8 or more
Average score School W	100.5	99.5	99.8	100.4	98.8	99.4	99.3	99.1
N	56	70	63	41	12	20	7	15
Average score School L	98.3	99.3	96.9	96.7	96.7	96.6	97.7	104.3
N	50	31	50	42	36	35	30	24

Table 6 reveals a consistently curvilinear relationship between cooperation on the Free Choice Test and size of family. The score increases, reaching its maximum in the 4-or-5-child group and then decreases uniformly. Those who come from homes of 4 or 5 children were most cooperative, "only" children and those coming from large families, 8 or more children, were least cooperative. The correlation ratio, η_{yx} was $.163 \pm .02$ for the size of family and score on Free Choice Test.²

2. *The Efficiency Cooperation Test.* This test measures the relative amount of effort a child puts forth on a group project. It is based on the difference in speed of work on a monotonous test when working for a personal prize and the speed of work for a group prize. As no names are written on the group papers, the child is free to work at any given speed while working for his group. The scores ranged from high negative, when work for self was at a higher speed, to positive, when work for group was at a higher speed. The correlation between this score and size of family differed

²See (8).

for the three schools. In School L the correlation was positive, $.1023 \pm .04$, in School E it was practically zero, while in School W the correlation was negative, $-.1633 \pm .03$.

Table 7 presents the average scores by the number of children in the family for Schools W and L. This table reveals that, while in School W the children of very large families (8 or more children) score lowest, in School L they score highest. The "only"-child group scored highest in School W and only average in School L. The relationship between size of family and score on this test is less consistent than with score on the Free Choice Test.

3. *A vote involving helpfulness.* In connection with a contest between different classrooms, money prizes were to be given to the winning groups. The children of each classroom were then allowed to vote on how the prize, if won, should be spent. The following vote form, devised by the Character Education Inquiry, was used.

"Write your name on the slip and then mark 1 what you would most like to do with the money. Mark 2 your second choice. Mark 3 your third choice. Mark 4 your fourth choice. Mark 5 your fifth choice.

- (a).....Give all the money to the boy or girl scoring highest in the test.
- (b).....Buy something for our school, such as bats, balls, skipping-ropes, a picture.
- (c).....Buy something for the room, such as a picture, a globe or gold fish, a plant.
- (d).....Divide the money equally among the members of the class.
- (e).....Buy something for some hospital child or some family needing help, or some other philanthropy.

The five possible choices were rated by a number of educators regarding the degree of helpfulness or charitableness they display. The order of preference was *e, b, c, a, d*. A child who marked the choices in this order scored 10. Any other vote was scored on the basis of the deviation from the educators' order of preference.

The correlations were then computed between this score and the number of children in the family for the children of Schools E and L. The correlations were both negative. In School E it was $-.1080 \pm .04$ and in School L it was $-.1059 \pm .04$. For the two schools combined the correlation was $-.1349 \pm .03$.

Table 8 gives the average score for five groups classified according to the number of children in the family. The figures reveal that the highest score on the vote was made by children coming from

families of 2 or 3 children, while the lowest score was made by those of large families. The relationship, however, was curvilinear. The score on the vote increased and then decreased with the size of family. The correlation ratio, eta, of score and size of family for the two schools combined was $.239 \pm .03$.

TABLE 8
HELPFULNESS SCORE BY SIZE OF FAMILY

Number of children in family	School E		School L		Total	
	N	Score	N	Score	N	Score
1	48	6.0	40	5.3	88	5.6
2	51	6.0	33	5.7	84	5.8
3	61	6.6	48	5.4	109	6.0
4 or 5	48	6.0	82	5.0	130	5.0
6 or more	13	4.2	86	4.9	101	4.8

4. *Composite score of service and size of family.* The records of the Character Education Inquiry contained five different scores of service. In addition to the three tests of cooperativeness mentioned above, there was a measure of a child's charitableness, his readiness to contribute to charity such articles as pencil sharpener, drinking cup, etc. These articles were given to each child as a present from a friend of the school. An appeal was then made to send away the articles to poor children who do not possess them. A record was kept of the exact amount given away by each child. Appropriate score values were ascribed to each article.

Another record was obtained of a child's willingness to collect and paste pictures and jokes which were to be sent later to children in hospitals. The contribution of each child to this project was recorded and a score assigned.³

The above five measures of a child's serviceability were then combined into one composite score. Table 9 presents the average scores by size of family for Schools E, W, and L.

The correlation between this score and the number of children in the family was $-.10 \pm .02$ ($N=787$) for the three schools combined. The correlation ratio, eta, of this score with size of family was $.19 \pm .02$. The relationship between the two variables was curvilinear. This is inferred from the difference between the product-

³For description of these techniques see (5).

moment correlation and the correlation ratio, and from the figures in Table 9. These indicate that the relationship between the children's composite score of serviceability and size of family is similar to the relationship between the latter and the various scores of cooperation and helpfulness, the score increases and then decreases with size of family. The highest service score was made by children coming from homes of two children followed by those who come from families of three children; then comes the "only"-child group. Those of large families (5 or more children) were less cooperative.

TABLE 9
COMPOSITE SERVICE SCORE AND SIZE OF FAMILY

Number of children in family	1	2	3	4	5 or 6	7 or 8	9 or more
Service score	112.0	115.3	113.1	110.8	109.9	109.8	106.9
<i>N</i>	50	154	167	113	127	50	26

5. *Rating on cooperativeness and size of family.* Each child was rated on cooperativeness by classmates and teachers. Three such rating scales were devised and used by the Character Education Inquiry.

Rating by classmates. Ten word-pictures describing certain types of children are presented to each child. He is to read it and write after each description the name or names of those whom the description fits. For example:

- Here is someone who is ready to play or work with the rest even when he (or she) can't have his own way.
- Here is a very selfish person who will not spare anything. He always wants to know what he can get for himself and never thinks of others.

Half of the descriptions were of positive and half of negative character with respect to helpfulness and cooperativeness.

Teacher's rating. Teachers were presented with lists of words some describing a cooperative nature, kind, generous, sociable, charitable, etc., and others describing an uncooperative nature such as selfish, cruel, greedy, callous, etc.

For each child of her group a teacher was to check those words which described the child best.

On another rating scale five types of children were described. For example:

Cooperation:

Works with others if asked to do so.

Works better alone, cannot get along with others.

Works well and gladly with others.

Indifferent as to whether or not he works with others.

Usually antagonistic or destructive to joint effort.

The teachers were to indicate each child's place on this scale.⁴

The three ratings all of which pertained to the matter of cooperativeness were transmuted into standard scores and combined into one rating on cooperativeness. The rating scores ranged from -44 to +40. The product-moment correlation between this rating on cooperation and the number of children in the family was $.1003 \pm .04$. The teachers and classmates (of School W) tend to rate as more cooperative those children coming from larger families.

An inspection of the scattergram showed that the correlation surface was not linear. The correlation ratio, η , was therefore computed and was found to be $.2149 \pm .04$.

Table 10 presents the average ratings by size of family for the children of School W.

TABLE 10
RATING ON COOPERATIVENESS AND SIZE OF FAMILY

Number of children in family	1	2	3	4	5	6 or more
Average rating	2.25	4.95	5.85	7.17	6.00	5.40
N	54	71	61	41	12	34

The "only"-child group was rated as definitely least cooperative. The average rating increased and reached the maximum for those who come from homes of four children. The rating decreased then uniformly.

Both the objective tests, as well as the ratings on cooperativeness by classmates and teachers, showed a curvilinear relationship with size of family. In either case the score increased and then decreased with the number of children in the family.

⁴See (9).

The "only" children, although they were not the most cooperative, score relatively higher on the objective tests of cooperativeness than on the ratings in cooperativeness. It probably indicates the presence of a prejudice against them.

The maximum score on the ratings was obtained by a larger-family group (the 4-children group) than the group obtaining the maximum score on the objective tests (the 2-children group).

Ratings on cooperation show a greater positive correspondence to size of family than do scores on objective tests of cooperation. On the other hand, the scores on cooperation tests show a greater negative correspondence to size of family. In either case, however, the relationship is curvilinear, having both a positive and a negative trend.

Children of extremely small families (1 child) as well as those of extremely large families (6 or more children) were found less cooperative in objective tests of cooperation as well as on rating on cooperation than children who come from families of 2-5 children.⁵

A similar relationship is reported between a child's cooperation and popularity. Those who had no friends as well as those who had a very large number of friends were less cooperative than those of average popularity.

Persistence and Size of Family. The total persistence score used by the Character Education Inquiry was composed of two measures, a child's persistence in working at top speed on monotonous tasks, simple additions, resisting fatigue, and his persistence in reading a story printed in run-in type. Table 11 presents the average scores in persistence by size of family. This table shows that in each one of the three schools the children coming from large families (7 or more children) scored highest in persistence. The group of "only" children, however, did not score lowest in persistence in any of the schools. The correlation between persistence and size of family was positive but extremely low in each of the schools. For the three schools combined the correlation was $+.074 \pm .02$, and the correlation ratio was $.1219 \pm .02$. The relationship with persistence was curvilinear but somewhat in the opposite direction of what was found in cooperativeness. There, it will be remembered, the children of very small as well as those of very large families scored lower (were less cooperative) than the others. Here they scored higher (more persistent) than the other groups.

⁵See (8, Chap. X).

TABLE 11
PERSISTENCE SCORES AND SIZE OF FAMILY

Number of children in family	School E		School W		School L	
	N	Score	N	Score	N	Score
1	42	168.9	46	131.4	36	180.0
2	40	138.7	69	141.3	65	186.5
3 or 4	73	145.9	96	140.0	62	188.1
5 or 6	16	167.7	26	109.1	73	175.8
7 or more	7	220.0	40	160.0	21	210.2

Inhibition and Size of Family. The score of inhibition used by the Character Education Inquiry was based on tests of the child's ability to postpone the desire to manipulate a certain toy which was placed near but which he was not to touch before another unit of work was completed. The correlation between the total score of inhibition and the number of children in the family was $-.1362 \pm .02$ for Schools E, W, and L. The correlation ratio of inhibition to size of family was $.230 \pm .02$.

TABLE 12
INHIBITION SCORES AND SIZE OF FAMILY

Number of children in family	1	2	3	4	5 or 6	7 or 8	9 or more
Average inhibition	12.2	12.2	12.6	12.7	11.1	10.6	10.5
N	112	133	149	103	107	46	24

It will be noted from Table 12 that the inhibition score increased slightly and then decreased markedly with size of family. Those coming from homes of three or four children scored highest, followed by those of one-or-two-children families, while those of large families (5 or more children) scored lowest.

Population Differences. The above comparisons were based on the combined three schools. The schools differed from one another in the relationship between inhibition and size of family. In School L the correlation was markedly negative $-.2346 \pm .04$, while in Schools E and W the product-moment correlation was low but positive, $.1141 \pm .04$ and $.0206 \pm .04$, respectively. The diverse relationship is shown in Table 13. This table reveals distinct differences between

schools. In School L there is a consistent decrease in score with size of family. The "one"-child group scored highest, and those of large families scored lowest. In Schools E and W the children of large families scored highest, while the "only"-child group scored next to the lowest.

In addition to the Toy Inhibition Test, the Character Education Inquiry used a Picture Inhibition Test. It measured a child's ability to continue working on simple additions without being distracted by pictures and jokes distributed throughout the test blank. The two scores were then added into one composite score of inhibition.

TABLE 13
INHIBITION AND SIZE OF FAMILY BY SCHOOLS

Number of children in family	School E		School W		School L	
	N	Score	N	Score	N	Score
1	33	12.5	50	13.1	29	13.6
2	43	12.6	63	13.3	27	12.8
3	50	13.1	62	13.4	38	12.6
4	26	14.2	37	14.1	40	12.3
5 or 6	22	11.6	25	13.0	60	10.6
7 or more	8	14.5	55	14.2	49	9.6

TABLE 14
COMPOSITE INHIBITION SCORE AND SIZE OF FAMILY

Number of children in family	School E		School W		School L	
	N	Score	N	Score	N	Score
1	33	22.3	46	22.2	27	22.5
2	41	22.9	64	22.3	33	22.8
3	43	23.0	61	22.7	39	22.9
4	12	25.0	35	23.5	33	23.5
5 or more	30	22.8	34	22.8	100	20.9

Table 14 reveals that in each of the three schools the inhibition score increased, reaching the highest average in the 4-child family, and then decreased for those of large families (5 children or more).

Honesty and Size of Family. The honesty scores utilized in this study were all based on behavior involving honesty, and not on children's ideas about honesty. The honesty tests of the Character Education Inquiry present opportunities for honest or dishonest re-

actions under controlled conditions. Two composite honesty scores were obtained: (a) an objective measure of a child's honesty in various types of school work, and (b) an objective measure of a child's honesty in work done at home.

The composite score of honesty in school work. This composite score was based on a series of tests in which a child could raise his score in a school test by copying illegally from a key, by continuing to work after time was called, and by peeping when a task was to be done with eyes shut. The various scores were converted into standard deviation units and added into one composite total. This total score, the measure of a child's deceit in school work, ranged from -33, extremely dishonest, to +5, definitely honest. This is shown in Chart 2 which gives the distribution of honesty scores by the number of children in the family for 636 cases.

CHART 2
HONESTY SCORES BY SIZE OF FAMILY

	111	134	149	95	53	37	31	8	8	3	2	1	1	633
Honesty scores	+ 5			1	2									3
	+ 3	1	1	2	1									5
	+ 1	6	7	13	4	2	1							33
	- 1	11	6	11	8	3	2	1	2	1	1			46
	- 3	11	15	16	10	2	1							55
	- 5	11	16	15	11	4	1	1						59
	- 7	11	17	13	4	5	5	3		1				59
	- 9	8	11	6	14	3	3	2	1					48
	-11	7	12	11	5	8	1	1	3	1	1	1		51
	-13	10	12	10	9	7	3	5						56
	-15	13	12	11	4	4	6	2	1	1				54
	-17	9	8	9	7	3	3	4	1					44
	-19	2	10	8	4		5	2	1					32
	-21	4	4	12	2	3		4		1				30
	-23			4	3	2	1	2		1				13
	-25	2	2	2	4	3	4	1		1				19
	-27	3		3	2	2						1		11
	-29	1	1	2	1		1		1					7
	-31	1				1		3						5
	-33					1	1						1	3
		1	2	3	4	5	6	7	8	9	10	11	12	
		Number of children in family												

The correlation between deceit and size of family for schools E, W, and L was $.10 \pm .02$, and the correlation ratio, eta, was $.14 \pm .02$. This tendency of an increase in deception with the number of children in the family was found in each one of the three school

populations. The magnitude of the correlation, however, differed from school to school. The correlations were as follows for schools E, W, and L:

In School E it was	.0818 \pm .04
In School W it was	.0466 \pm .04
In School L it was	.2022 \pm .04

We note that in School L, the under-privileged population, the correlation between deceit and number of children in family is most pronounced. The children of the larger families of this population were more deceitful than those coming from smaller families. In the other two schools the negative relationship between honesty and size of family was less pronounced.

Table 15 presents the deception averages for each group classified according to the number of brothers and sisters.

TABLE 15
AVERAGE HONESTY SCORES BY SIZE OF FAMILY

Number of children in family	1	2	3 or 4	5 or 6	7 or 8	9 or 10	above 10
Deception	—9.97	—9.67	—10.08	—13.61	—15.31	—12.81	—16.70
<i>N</i>	111	134	244	90	39	11	7

It reveals that those who come from families of one ("only" child) or two children were the most honest (least deceptive) and that with the exception of one group, 9 or 10 children, the deception score increased consistently with increase in the number of siblings.

Honesty in home work. One of the Character Education Inquiry tests involved honesty in doing home work. The children were given units of work which were to be done at home without the help of any person or book. Their actual ability on those units were later tested by an equivalent form administered in school. The difference was apparently a measure of illegal help accepted in doing the home work.

The correlation between dishonesty at home and size of family was negative but extremely low, $-.047 \pm .02$. This negative relationship is probably due to the fact that in homes of smaller families the parents took more interest in the children's work and thus gave them help. In fact the "only"-child group scored lower than any

other group on this test of honesty, having the highest deception score in each one of our populations. It is probably the greater opportunity for receiving help rather than a greater tendency to be deceptive that lowered the honesty score of this group on the home-work honesty test.

Intelligence of Parents and the Number of their Children. A short test of intelligence and a test of moral knowledge were administered by the Character Education Inquiry to the parents of 218 children included in this study. The parents who submitted themselves to the test were probably the most intelligent in the group.

None of the parents of the children of School L have taken these tests. Those for whom these records were available come from our "average" population, School W. The relationship that was found is thus not representative of the total population.

TABLE 16
INTELLIGENCE OF PARENTS AND SIZE OF FAMILY

Number of children in family	1	2	3	4	5 or more
Father's score of intelligence	80.0	86.8	86.7	93.3	76.0
Mother's score of intelligence	77.3	86.0	86.4	88.0	78.0
Mid score of par- ents' intelligence	78.7	86.4	86.6	90.7	77.0
<i>N</i>	16	19	23	4	3

Table 16 reveals a curvilinear relationship between parental intelligence and size of family. The intelligence score increased from those having one to those having four children and then dropped considerably. As was pointed out previously, selective factors are probably responsible for the apparent differences. It is likely that among those who took the test were the poorer element of the one-child family (some of them being too busy to have spent time on it) and the best of the larger (2-4 children) families, the poorest of the latter being unable to comprehend the directions.

The parents of the largest families (5 or more children) were of somewhat lower intelligence than the others. Among the one-to-

four-child families the intelligence increased with the number of children.

It should be noted that among the one-to-four-child families the father was of higher intelligence, while among those of 5 or more children the mother was of higher intelligence. The number, however, was too small to warrant any generalizations.

TABLE 17
MORAL KNOWLEDGE OF PARENTS AND SIZE OF FAMILY

Number of children in family	1	2	3 or more
Father's score in moral knowledge	154.4	149.0	146.6
Mother's score	141.7	150.0	148.0
Mid-score of parents	148.1	149.5	147.3
<i>N</i>	16	19	30

Table 17 indicates that the father's moral knowledge score decreases consistently with the size of family. The mother's moral knowledge score, however, is highest among the two-children families. It is interesting to note that of the one-child families the fathers were of the highest moral knowledge, while the mothers were of lowest moral knowledge scores. In moral knowledge there is a distinct difference between father and mother. The moral knowledge of the father decreased consistently with increase in size of family. In the case of the mother there was a marked increase in moral knowledge from the one- to the two-child family, followed by a slight decrease in the larger families (3 children or more).

Rating on Character and Size of Family. Each child was rated by his or her teacher on such aspects of personality as cooperation, persistence, inhibition, reliability, honesty, etc. Each trait was described in five steps ranging from the extremely negative, through zero, to the extremely positive aspect of the trait. The teacher was to indicate where the child belonged on the scale of each trait. The ratings were converted into standard units and added into one composite character rating. Table 18 gives the average ratings for each group according to the number of children in the family.

TABLE 18
CHARACTER RATINGS BY SIZE OF FAMILY

Number of children	1	2	3	4	5 or more
Rating average	66.0	73.6	76.6	72.1	69.7
N	52	70	63	39	44

Table 18 reveals a consistent but curvilinear relationship between children's ratings on character and the size of their families. The rating is lowest for the "only"-child group. It increases to a maximum in the three-child family and then decreases.

In the opinion of teachers (of School W), then, the most superior character is possessed by those children who have two brothers or sisters (3 children in the family), followed by those of 2 or those of 4 children in the family, then comes those of the large families (5 or more children), and last come the "only" children.

The "only" child is usually considered a spoiled child and probably has greater difficulties in adjustment with the teacher than the other children. He was thus rated frequently as selfish, uncooperative, low in persistence, honesty, and inhibition.

SUMMARY

The size of family correlates negatively with intelligence, moral knowledge, cultural background, and with honest behavior. The correlations, though showing a negative trend, do not reveal a linear relationship between the above factors of personality and size of family. The "only"-child group is not superior to the others (except in moral knowledge). The highest intelligence scores are made by those coming from homes of two children.

The size of family shows a definitely curvilinear relationship with cooperativeness and helpfulness. Those of very small families as well as those of very large ones scored lower than those of average size families. There was a similarly curvilinear relationship between size of family and the children's scores on tests of inhibition.

Size of family correlates positively with scores of persistence, children of large families scoring highest on persistence tests.

Ratings by teachers and classmates show a curvilinear relationship to size of family. The rating increases and reaches its maximum for the three- or four-child family and then decreases.

The correlation between size of family and intelligence and moral knowledge of parents cannot be determined on our population because the few parents for whom test scores were available were not representative of the total population. There was, however, a markedly negative relationship between the social status of the parents and the number of their children.

If we consider the families tri-serially, those having one child, those having two to five children, and those having six or more children, we find the following relationships.

The "only" child is above average in intelligence, in moral knowledge, cultural background, and honesty. He is just average in cooperativeness, and persistence; and below average in inhibition and ratings by teacher and classmates.

The children of small families are highest in intelligence, in honesty, in inhibition, and in ratings by teacher and classmates.

The children of large families are lowest in intelligence, moral knowledge, cultural background, honesty, cooperativeness, inhibition, parents' intelligence, and moral knowledge. They are highest in scores of persistence.

REFERENCES

1. CHAPMAN, J. C., & WIGGINS, D. M. Relation of family size to intelligence of offspring and socio-economic status of family. *Ped. Sem.*, 1925, **32**, 414-421.
2. DAYTON, N. A., Order of birth and size of family. *Amer. J. Psychiat.*, 1929, **8**, 979-1006.
3. DUFF, J. F., & THOMSON, G. H. The social and geographical distribution of intelligence in Northumberland. *Brit. J. Psychol.*, 1923, **14**, 192-198.
4. HAGGERTY, M. E., & NASH, H. B. Mental capacity of children and paternal occupation. *J. Educ. Psychol.*, 1924, **15**, 559-572.
5. HARTSHORNE, H., MAY, M. A., & MALLER, J. B. Studies in service and self-control. New York: Macmillan, 1929. Pp. xxiv+560.
6. LENTZ, T., JR. Relation of I.Q. to size of family. *J. Educ. Psychol.*, 1927, **18**, 486-496.
7. MACDONALD, H. The social distribution of intelligence in the Isle of Wight. *Brit. J. Psychol.*, 1925, **16**, 123-129.
8. MALLER, J. B. Cooperation and competition. *Teach. Coll. Contrib. Educ.*, 1929, No. 384. Pp. 176. (Esp. Chap X, "The environment.")
9. MAY, M. A., & HARTSHORNE, H. Recent improvements in devices for rating character. *J. Soc. Psychol.*, 1930, **1**, 66-77.
10. PEARSON, K., & MOUL, M. The problem of alien immigration into Great Britain. *Ann. Eug.*, 1925, **1**.
11. SANDIFORD, P. Paternal occupations and intelligence of offspring. *School & Soc.*, 1925, **13**, 117-119.

12. SLAWSON, J. The delinquent boy. Boston: Badger, 1926. Pp. 485.
13. SUTHERLAND, H. E. G. Relation between I.Q. and size of family. *J. Educ. Psychol.*, 1929, **20**, 81-90.
14. SUTHERLAND, H. E. G., & THOMSON, G. H. The correlation between intelligence and size of family. *Brit. J. Psychol.*, 1926, **17**, 81-92.
15. Terman, L. M., et al. Genetic studies of genius. Vol. I. The mental and physical traits of a thousand gifted children. Stanford University, Calif.: Stanford Univ. Press, 1925. Pp. xvi+648.
16. THURSTONE, L. L., & JENKINS, R. L. Birth order and intelligence. *J. Educ. Psychol.*, 1929, **20**, 641-651.
17. WINSTON, S. R. The relation of certain social factors to fertility. *Amer. J. Sociol.*, 1930, **35**, 753-764.

*Teachers College
Columbia University*

LA GRANDEUR DE LA FAMILLE ET LA PERSONNALITÉ DES ENFANTS

(Résumé)

L'Enquête sur l'Enseignement de Caractère ("Character Education Inquiry") a testé un groupe de 802 enfants au point de vue de l'intelligence, l'honnêteté, l'esprit de coopération, et la maîtrise de soi. Leurs maîtres et leurs camarades d'école les ont évalués aussi sur plusieurs échelles d'évaluation de caractère. Cette enquête est une étude des rapports entre chacun des aspects de personnalité cités en-dessus et le nombre d'enfants dans la famille.

La population étudiée s'est composée d'enfants âgés de 9 à 18 ans, dans les "Grades" V à VIII. Ils ont été élèves de trois écoles, représentant en nombres approximativement égaux les niveaux socio-économiques inférieur, moyen, et supérieur.

(Le nombre d'enfants dans chaque famille a varié entre 1 et 10. Aussi a-t-on divisé les enfants en dix groupes. On a déterminé la corrélation entre chaque variable et la grandeur de la famille. On a computed le résultat moyen de chaque test et de chaque évaluation pour chacun des dix groupes. On a fait une attention spéciale aux résultats de personnalité du groupe des "enfants uniques" et aux résultats des enfants des plus grandes familles.)

Dans plusieurs cas on a testé aussi les parents des enfants au point de vue de l'intelligence et de la connaissance de la moralité. On a déterminé la corrélation entre les résultats des parents et le nombre de leurs enfants.

On donne les résultats en tables et tableaux. Ils montrent une corrélation généralement négative entre la grandeur de la famille et des aspects de personnalité tels que l'intelligence, connaissance de la morale, fond de culture, l'état social, et l'honnêteté. On a trouvé une corrélation positive entre la grandeur de la famille et les résultats des tests de persistance.

On a trouvé une corrélation curviligne avec les résultats de l'esprit de coopération et ceux de caractère. Les enfants des très petites familles, ainsi que ceux des très grandes, ont montré moins de coopération et n'ont pas reçu une évaluation aussi élevée que celle des enfants des familles de grandeur moyenne.

Le groupe des "enfants uniques" a obtenu un résultat très élevé dans la connaissance de la morale, un résultat un peu au-dessus de la moyenne dans l'intelligence, un résultat moyen dans l'honnêteté, l'inhibition et l'esprit de coopération, et un résultat au-dessous de la moyenne dans la persistance. Leurs maîtres et leurs camarades les ont évalués au-dessous de la moyenne au point de vue de caractère.

Les enfants des très grandes familles ont reçu un résultat moins élevé que les autres groupes dans chacun des tests cités en-dessus, à l'exception du test de persistance. Leurs maîtres et leurs camarades les ont évalués au-dessous de la moyenne.

(Le rapport entre la grandeur de la famille et la personnalité a été constant dans sa direction pour les trois niveaux sociaux mais la magnitude de la corrélation a été différente dans les diverses écoles. Elle a été la plus frappante dans l'école du niveau social le moins élevé, et la moins frappante dans l'école du niveau social moyen.

MALLER

DIE GRÖSSE DER FAMILIE IN IHREM VERHÄLTNISS ZUR PERSÖNLICHKEIT DER NACHKOMMENSCHAFT

(Referat)

Es wurde eine Gruppe bestehend aus 802 Kindern geprüft durch die 'Forschung über die Charakterbildung' ('Character Education Inquiry') in Bezug auf Intelligenz, Ehrlichkeit, Fähigkeit zur Mitwirkung, und Selbstbeherrschung. Die Kinder wurden ebenfalls von Lehrern und Mitschülern nach mehreren Massstäben zur rangmässigen Klassierung in Bezug auf Charaktereigenschaften ('Character Rating Scales') zensiert. Diese Untersuchung betrifft die Beziehung jeder der oben erwähnten Aspekte der Persönlichkeit zur Zahl der Kinder in der Familie.

Die untersuchte Gruppe bestand aus Kindern im Lebensalter von 9 bis 18 Jahren aus der Vten bis durch die VIIIe Schulklasse. Sie kamen aus drei Schulen worin respektiv, in ungefähr gleichem Masse, die unteren, mittleren, und oberen sozial-ökonomischen Schichten vertreten waren.

Die Zahl der Kinder in der Familie erstreckte sich von 1 bis 10. Die Kinder wurden folglich in zehn Gruppen geordnet. Man berechnete die Korrelation jeder Variable mit der Grösse der Familie. Die Durchschnittszahl ('average score') für jede Prüfung und jede Zensierung wurde für jede der zehn Gruppen ermittelt. Man lenkte die Aufmerksamkeit besonders auf die 'Persönlichkeitsziffern' ('personality scores') der Gruppe der 'Einzelkinder' und ebenfalls der Gruppe der Kinder aus sehr grossen Familien.

In mehreren Fällen prüfte man auch die Eltern der Kinder in Bezug auf Intelligenz und moralische Kenntnisse. Man berechnete die Korrelation der Ziffern ('scores') der Eltern mit der Zahl ihrer Kinder.

Die Ergebnisse werden auf Tabellen und Karten dargeboten. Sie zeigen eine gewöhnlich negative Korrelation zwischen Familiengrösse einerseits und verschiedenen Aspekten der Persönlichkeit wie zum Beispiel Intelligenz, moralische Kenntnisse ('moral knowledge'), kulturelle Umgebung, sozialer Stand, und Ehrlichkeit andererseits. Es zeigte sich eine positive Korrelation zwischen Familiengrösse und den in Prüfungen der Ausdauer ('persistence') erzielten zahlen. Die Korrelationen mit Mitwirkungsfähigkeit ('cooperativeness') und mit den Charakterzensierungen ('character ratings') waren

bogelineal ('curvilinear'). Die Kinder aus sehr kleinen und auch die aus sehr grossen Familien waren weniger mitwirkungsfähig und erhielten niederigere Zensuren wie die Kinder aus Familien von mittelmässiger Grösse.

Die Gruppe der 'einzelnen' Kinder zeigte sich in moralischen Kenntnissen am besten, in Ehrlichkeit, selbst-Beherrschung, und Mitwirkungsfähigkeit als mittelmässig, und in Ausdauer als unter dem Durchschnitt stehend. In den Charakterzensurierungen durch Lehrer und Mitschüler stand sie unter dem Durchschnitt.

Das Verhältniss zwischen Familiengrösse und Persönlichkeit war in seiner Richtung für die drei sozialen Schichten konsequent; die Höhe der Korrelation war aber in den verschiedenen Schulen verschieden. Sie war in der Schule aus der untersten sozialen Schicht am meisten und in der Schule aus der mittleren sozialen Schicht am wenigsten ausgeprägt.

MALLER

IS PERSEVERATION A FUNCTIONAL UNIT PARTICIPATING IN ALL BEHAVIOR PROCESSES?*

From the Psychological Laboratory of the University of Oregon

HERBERT H. JASPER

Perseveration may be defined as the tendency of a set of neurons, once excited, to persist in the state of excitation autonomously, showing resistance to any change in this state. This quality of the nervous process, which appears analogous to the inertia of physical matter, has been observed in the sensory, motor, and ideational fields of behavior. A critical investigation of the functional unity of the perseverative tendency in various behavior processes is the object of this study.

Aristotle, in his chapter on memory and recollection, makes some interesting observations regarding the recall process in different types of individuals. He states that:

"Many persons are made very restless when they cannot recall a thing, and when quite inhibiting their thought, and no longer trying to remember, they do recollect nevertheless, as is especially true of the melancholic. For such persons are most moved by images.....Especially disturbed are such persons as have moisture about the region of sensation; for they do not easily come to rest after being stirred into motion, until they attain the thing sought for, or the movement has taken its proper course. Consequently, the feelings of anger and fear, when they once set up a movement, do not cease although opposing movements are started against them, but on the contrary persist towards their own aim. This affection resembles names, melodies, and words, when they are given violent utterance. For after one has ceased, the singing or speaking recurs involuntarily. Further, those whose upper body is too large, and also dwarfish persons, have less power of recollection than those of the opposite physical structure, because the former are too heavy about the organs of sensation, and because the initial movements cannot persist but are destroyed, and direct movement in

*Recommended by Edmund S. Conklin and received for publication by Carl Murchison of the Editorial Board, January 15, 1930.

¹The helpful suggestions of Professor William Griffith of Reed College, Portland, Oregon, Professors Edmund S. Conklin, Howard R. Taylor, and Robert H. Seashore of the University of Oregon have been greatly appreciated from the initiation to the completion of this study.

the process of recollection cannot readily take place." (*Psychology*, Hammond's translation, p. 211-212.)

This is the first record we have of the observation of perseveration in the neural process and of the relationship of perseveration to emotional and bodily types.

In 1894 the term "perseveration" was used by Neisser (20) to indicate abnormally persistent repetition of an activity after the activity normally would have been completed. This term was originally used as descriptive of a particular symptom to be recognized in the psychiatric clinic.

G. E. Müller (27), in 1900, and Foster (9), in 1914, observed perseverative phenomena in their work on memory. Some individuals, more than others, would respond incorrectly in an attempt to recall a series of syllables recently learned but would respond with syllables that were correct for a series of syllables learned previously. Müller called this phenomena "perseveration," while Foster attempted to explain it on the basis of association.

In 1902 the psychiatrist Otto Gross (11) introduced the term "secondary function of the nervous system" for the term "perseveration." The "primary function" was the initial activity of a stimulated sensory-motor process during stimulation. The "secondary function" was the tendency of this process to continue spontaneously in the state of excitation for a certain period of time after the cessation of the stimulus. Gross gives some evidence in support of Aristotle's association of this function with melancholics and asthenics. Wiersma (36), in 1906, claimed to be able to distinguish between "mania, melancholic-paranoia, and normals" by measuring the degree of "secondary function" in these three types of individuals. He used a measure of the lag or immediate after-effect of sensation as indicative of the degree of "secondary function."² The degree of "secondary function" seemed to increase along the line manic-normal-melancholic. These results are not corroborated in full by the more recent study of Wynn Jones (21).

The works of Heymans and Brugmans (15) in 1913, Wynn Jones (19) in 1915, Lankes (24) in 1915, Bernstein (3) in 1924, Hargreaves (13) in 1927, and Wynn Jones (20, 21) in 1928 have

²Wiersma's measures were (a) flicker fusion speed of rotation, (b) the time of adaptation to the liminal light threshold, and (c) the time required to return to normal sensitivity to a weak electric shock after stimulation by a strong shock.

added a number of measures of perseveration in the sensory, "ideational," and motor fields. These measures may be listed as follows:

MEASURES OF PERSEVERATION IN THE SENSORY FIELD

1. *Flicker Fusion Speed of Rotation.* The speed of rotation necessary to produce gray from a two-sector color disc was first employed by Wiersma and Heymans and Brugmans. Lankes added some technical improvements which were used by Jones. Heymans and Brugmans used the speed of rotation necessary to see a continuous light through a slit in a revolving disc as an additional measure.

2. *Light Adaptation.* Wiersma used the time necessary to adapt to the liminal threshold, while Heymans and Brugmans and Jones (20) used a dim light of constant intensity instead of the liminal threshold.

3. *Adaptation to Electric Stimulation.* Wiersma (36).

4. *Adaptation to Sound Stimulation.* Heymans and Brugmans (15).

MEASURES OF PERSEVERATION IN THE "IDEATIONAL" FIELD³

1. *Recall Tests.* Müller and Foster used syllables as described above. Lankes used narratives. A passage of prose was read three or four times to the subject and he was quizzed on it. Then another passage was read and the responses to questions on the second passage were indicative of the perseveration of the first passage.

2. *Towns Test.* Instructions were given to write as many towns as possible beginning with a certain letter. Rapidity of association was to indicate a lack of perseveration. (Hargreaves)

3. *Word-Building.* As many words as possible are to be made out of the letters of four words given. One minute is allowed for each word and one mark given for each word made. A low score would indicate perseveration. (Hargreaves)

4. *Nouns Test.* Instructions are given to think of as many nouns as possible beginning with a certain letter (b and s used). [Jones (20)]

5. *Animals Test.* One mark is given for each name thought of in one minute. [Jones (20)]

6. *Whipple's Ink Blots.* One mark is given for each meaning. [Jones (20)]

³These measures might be considered in the motor field, depending upon your interpretation of "ideational."

7. *Essays*. The ease of writing essays in a given time. (Lankes)
8. *Associative Reaction*. The repetition of identical associations to a series of stimulus-words in which some words were repeated. (Lankes)

MEASURES OF PERSEVERATION IN THE MOTOR FIELD

1. *Reverse Order Letters*. Five letters were first written 40 times in one order and then 20 times in the reverse order. The difficulty in shifting from the first to the second order was to indicate perseveration (Heymans and Brugmans op. cit., Lankes op. cit., Bernstein op cit.)

2. *Letter Cancellation*. The subjects are given a period of practice in cancelling all the "a"s and "t"s, e.g., in a series of letters. The persistence of the tendency to cancel all the "a"s and "t"s in a succeeding series in which all the "p"s and "r"s are to be cancelled in the measure of perseveration. (Lankes op. cit.)

3. *Inverted S Test* (other letters used also). S is written repeatedly in the usual way and then as it would appear in a mirror. [Jones (20), Bernstein and Hargreaves]

4. *Tapping Test*. A slow optimal tapping-rate is thought to be indicative of perseveration. (Heymans and Brugmans, Lankes, Bernstein, and Hargreaves)

5. *Mirror-Drawing*. [Jones (20)]

6. *The "it" Test*. A passage of prose was written first as fast as possible being careful to dot all the "i"s and to cross all the "t"s and then it was written with the instruction not to dot the "i"s or to cross the "t"s. [Jones (19-21) Bernstein, and Hargreaves]

7. *Reverse Stroke Test*. A series of digits were written repeatedly in a certain order for one minute. They were then written with the reverse stroke for one minute, i.e., each digit was to be written in the reverse direction (beginning where you usually finish) and the series of digits were to be written in the reverse order. [Bernstein, Hargreaves, and Jones (20, 21)]

8. *Triangles Test*. The difficulty in shifting from drawing triangles with the apex upwards to drawing them with the apex downward was the measure of perseveration. (Bernstein)

9. *Capitals Test*. A passage of several sentences was written for a given time as given with several words and letters in capitals. The passage was then to be written making all the small letters capitals and the capitals small letters. (Bernstein)

10. *Vertical-Horizontal Test.* Four rows of simple geometric figures were given to be copied for one minute. The same figures were then to be written substituting a vertical line for a horizontal line and a horizontal line for a vertical line for two minutes. (Bernstein)

11. *The "ea" Test.* A given passage was written as given for two minutes and then written with the instructions to place an "a" after every "e" occurring in the passage. (Bernstein)

Lankes used an interrogatory on the perseverative tendency and Bernstein used teacher's ratings in addition to some of the above measures.

A survey of the above alleged measures of perseveration is indicative of the wide variety of responses which are thought to be included under the one concept of perseveration. Spearman (33) has thought that there is sufficient experimental evidence from the administration of the above measures (he does not include the work of Hargreaves) for the postulation of a functional unity in all these different manifestations of the perseverative tendency. Furthermore, he comes to the conclusion that the tendency toward introversion as proposed by Jung and the tendency toward "subjectivity" as proposed by Stern are essentially the same as the perseverative tendency or "general inertia." The individual whose neural processes tended to persist for some time after having once been aroused would, of necessity, have his attention directed upon these processes in consciousness and become introverted in consequence (33, Chapters IV and XVII).

Kelley comments on these conclusions of Spearman as follows (22, p. 10):

"Surely so vital an issue as this could well be made a matter of first importance, even if the investigator is studying merely the one or the other trait, and not merely trying to see things in proper perspective. If this latter is the object, then not only should introversion and perseveration be studied together, but also along with many other traits, of which some will probably be related to these two and some will be entirely independent. A comprehensive study of this last sort has never been made."

This suggestion has come into print after the completion of the experimental work in this study, but the general nature of this work might well be thought of as a preliminary attempt to carry out his suggestion at this point.

Spearman himself is not without suggestions in this regard. Speaking of perseveration he states in conclusion (33, p. 306):

"we have been examining the greatest of all faculties, if by this may be signified the one which has been most lavish of promises for individual psychology. It is also among the greatest—only second to 'intelligence'—in the sense that whilst all the other asserted faculties have proved to be baseless, this perseverance now shows itself to be at any rate a *half truth*. For there does appear to exist, as a unitarily functioning factor varying in degree from one individual to another, a tendency for mental processes to have a certain lag or inertia and in this meaning to 'perseverate.'"

and further (33, p. 307):

"Turning to the practical standpoint, the prospect here is extraordinarily hopeful. When once the pack of modern investigators can be called off the many false scents of illusory faculties to this genuine trail; when the perseverance, already measurable even by groups, has been evaluated for persons of diverse age, sex, character, and social status; when the connection has been traced out which it bears to success in different branches of education and varieties of vocation—then perhaps psychological science will have made a second advance not much less in magnitude than that which is being achieved with respect to 'intelligence.'"

In the present author's opinion the "genuineness" of the perseverance "trail" lies only in the realm of interesting hypotheses if the experimental evidence is given critical examination. The functional unity of perseverance itself has yet to be definitely established before any such extensive program of research as Spearman has outlined would be of significance.

THE MEASURES OF PERSEVERATION USED IN THIS INVESTIGATION

Many of the measures of perseverance used in this investigation were near duplications of those used in previous studies as will be recognized from the preceding survey. Reliabilities in terms of the product-moment r with the Brown-Spearman (prophecy formula) correction are given for those measures amenable to this procedure. Eighty cases were used in the calculation of the reliability coefficients.

Sensory Measures⁴

1. *Light Adaptation.* ($r_{11} = .78$, the second measure with the third measure) [See references (4) and (14) for the neurological basis of this reaction.] This measure was the time (in seconds) required to see a light of constant low illumination after being

⁴Measures of flicker-fusion speed of rotation were attempted but proved to be too unreliable for use in this investigation although methods used by former investigators were employed.

adapted to the illumination of a 100-watt Mazda flood-light. A double contact switch made it possible to turn off the flood-light and, at the same instant, start an electrically innervated stop-watch. The subject pressed a reaction key as soon as he was able to see three spots of light at very dim illumination. The three dim lights were arranged in brightness so that they were seen in sequence as adaptation progressed. The first light seen provided a fixation point for the appearance of the other two lights so that adaptation would be very nearly foveal for all the subjects. The three lights were 1 cm. apart at a distance of 150 cm. from the corneal apex. A report of the relative positions of the three lights served as a check on the reliability of the response. The light was diffused through different thicknesses of gray tissue paper to obtain the different degrees of brightness.

The mean of the second and third measures of this adaptation time was used as the perseveration measure.

2. *Mean Simple-Reaction-Time to Visual Stimuli.* ($r_{11} = .91$, even vs. odd measures with Brown-Spearman correction) A 1/100 sec. electrically innervated stop-watch with a sliding switch to start the stop-watch at the instant of the visual stimulus was used for the reaction-time measures. The mean of 20 measures was used. An interval of 10 minutes elapsed between each 10 measures. The precautions of Woodrow (34) and Johnson (18) were taken into consideration. The visual stimulus was a frosted 40-watt tungsten filament globe seen as reflected upon a white background through a hole in a cardboard screen of neutral grey.

3. *Mean Choice-Reaction-Time to Visual Stimuli.* ($r_{11} = .76$, even vs. odd with Brown-Spearman correction) Red, green, blue, yellow, and orange colored electric globes were enclosed behind a cardboard screen by a white cardboard reflector. In the center of the screen was a hole 4 cm. in diameter 150 cm. in front of the observer. Measures of choice-reaction-time to color were obtained in the usual way by the same apparatus as was described for simple-reaction-time. The mean of 20 measures was used; the measures being given in groups of 10.

4. *Sigma Simple-Reaction-Time to Visual Stimuli.* (Standard Deviation)

5. *Sigma Choice-Reaction-Time to Visual Stimuli.* (Standard Deviation) The theoretical assumption warranting the use of these measures of reaction-time is that the perseverator, although perhaps slower in reaction, would be less variable than the non-perseverator.

Motor Measures

6. *Tapping Rate.* (Three measures were taken five minutes apart: $r_{12} = .98$, $r_{23} = .99$, $r_{13} = .97$.) The subject was instructed to tap with a stylus on a metal plate at his most convenient rate of speed. The number of taps was recorded on an electric Hollerith counter. The score was the mean of three measures in the number of seconds per 100 taps.

7. *Star-Tracing in Mirror-Vision.* ($r_{11} = .88$, time plus errors the first tracing with time plus errors the second) The apparatus used for this measure was very similar to Snoddy's "Stabilimeter" (32) with an electrical recording of errors made. The subjects were paced to equalize seconds and errors in a few preliminary trials. The score for perseveration was seconds plus errors on the first tracing, providing a rough measure of initial efficiency in adaptation to a new task, i.e., the reversal of cues to eye-hand coordination in mirror-vision.

8. *Motor Inhibition (I).*⁵ An eye-finger coordination habit was well learned for pressing finger keys on a Stoelting five-finger keyboard in response to five different colored lights exposed by the method described for choice-reaction-time. Each light was exposed for the same length of time each series of exposures, but each of the five lights had a different duration of exposure under mechanical control. The Hollerith counter used for recording correct responses could not be innervated unless the correct key was pressed for a certain color during the time of exposure of that color, thus eliminating a learned rhythm of response and making possible an exact quantitative record of the correct responses made. The right hand was used with right-handed cases and the left hand with left-handed cases. The colors were exposed in the following order, the experimenter giving (verbally) the correct finger for each color as it appeared: Y-4 (4th finger), G-5, B-3, R-1, Orange-2, R-1, B-3, G-5;

⁵A complete description of the apparatus for Measures 8 and 9 will be omitted from this paper. Further research on motor inhibition is being carried on with this apparatus at present and is to be reported in the near future with a complete description of the apparatus used. For similar apparatus see bibliography references (25) and (29). From the nature of the response measured and the method of measurement the reliability of these measures should be very high. All of these measures were taken in 15 minutes and consisted of a total of 240 separate reactions. Other reactions of a very similar nature have been shown to have a high reliability (25, 30, 12).

then the series was begun over again with Y-4, continuing in this manner for 80 exposures. Then a series of 120 exposures was given with no coaching as to the correct finger for each color and a record of the number of correct responses was taken. This series will be called the simple-discrimination series (*Sd*). At the end of this period the habit was learned to the point of automaticity in most of the subjects. Immediately following this *Sd* series another series of 120 exposures was given exactly in the same manner as the *Sd* series with the exception that 18 of the colors in the series were dimmed appreciably. The observers were instructed *not* to respond to any color in this series (inhibition discrimination series, *Id*) which was dimmed. The subjects were conscious of errors when they responded to dim colors. The inhibition score, *I*, was the number of dim colors responded to in the *Id* series. (This record was taken on another Hollerith counter while the record of the correct responses to the bright colors in the *Id* series was being recorded by the first counter.) One would expect, from theoretical assumptions, that the perseverator would tend to respond as in the *Sd* series to the dim colors in the *Id* series, making a high inhibition score, *I*, indicate a high degree of perseveration.

9. *Motor Interference (Sd—Id)*. This was a measure of the amount of interference in the learned motor pattern described above caused by the inhibition process in the series *Id*. [For similar measures see Bair (1), Bergstrom (2), and Culler (6).] The score *Sd* was the number of colors responded to with the correct fingers in the series of exposures where all colors were at full brightness and no inhibition instructions were present. The score *Id* was the number of colors responded to with the correct fingers in the series in which 18 out of the 120 exposures were dimmed and instructions to inhibit any response to dimmed colors were given. Hence the score *Sd—Id* would indicate the amount of interference caused by the change in the situation of the learned motor pattern. It was supposed that with the perseverator the learned response in the series *Sd* would persist into the second series, *Id*, with little variation due to the slight change in the stimulus pattern and instructions of the series *Id*. In this case a small *Sd—Id* score should indicate a high degree of perseveration. This measure (as well as 8) might be vitiated somewhat by the difficulty the perseverator is supposed to have in adapting to a new situation, but the situations are so nearly

the same that one would not expect an appreciable effect from this factor.

*10. *Changed Order Letters*. The letters "l d g x p n" were written as rapidly as possible for one minute with marks placed at 15-second intervals. Then, at the signal "change" the letters were written in the order "g d x n l p" for one minute, marking 15-second intervals as before. The percentage of all the letters written in the second column that were written in the first 15-second interval was subtracted from the percentage of all the letters written in the first column during the last 15-second interval. This measure is somewhat similar to the interference score, $Sd - Id$.

*11. *Cancellation*. ($r_{11} = .39$, between two measures described below.) The Woodworth-Wells series of digits reproduced three times on one page was used for this measure. Instructions were given to cancel every 2 and 7 in the first set of digits. At the end of two minutes, instructions were given to cancel every 3 and 9 in the second set of digits and then, after two minutes, instructions were given to cancel every 8 and 6 in the third set of digits. Two measures of perseveration were thus obtained; (a) the difference between the efficiency in the first set and that in the second set, and (b) the difference between the efficiency in the second set and that in the third set. (This test was scored according to Whipple's *Manual of Physical and Mental Tests*, Vol. 1, p. 313.)

*12. *Arithmetic Shift (add to multiply)*. [The arithmetic shift tests were based on the work of Jersild (17).] Three pages of digits under ten were presented in pairs. On the first page the subjects were instructed to add each pair for $2\frac{1}{2}$ minutes. On the second page the subjects were to multiply each pair for $2\frac{1}{2}$ minutes. On the third page they were to add the first pair and multiply the second pair and so on alternately for 5 minutes. The difference between the number of computations made the first five minutes and the number made the second five minutes was used as the measure of perseveration. A large score should indicate perseveration.

*13. *Arithmetic Shift (subtract to divide)*. This measure was the same as number 12 with the exception that the numbers were subtracted and divided instead of added and multiplied. In this measure (and to some extent in number 12) the difficulties in computation could very easily have masked all perseverative effect.

14. *Narrative Memory Test (IR—Immediate Recall)*. A short narrative was read to the subjects and 30 questions were asked im-

mediately following the reading. The number of correct responses should be greater with the perseverative individual.

15. *Narrative Memory Test* (*IR-DR*—Immediate Recall—Delayed Recall). The same questions were asked again as in number 14 but after a period of seven days during which no review or rereading of the narrative was permitted and the subjects were not informed that a second recall was to be given. Supposedly the narrative would tend to persist for the perseverator; facilitating the delayed recall in comparison with the immediate recall. In 50% of the cases there was either no loss or an improvement in the number of correct responses in delayed recall. The scores (*IR-DR*) ranged from plus 10 to minus 14, indicating a wide individual difference in the perseveration of the narrative. The story was a very exciting, highly emotional type which undoubtedly contributed to the improved delayed recall in so many cases.

16. *Questionnaire*. ($r_{11} = .59$, $P.E. = .06$, even *vs.* odd with Brown-Spearman correction.) This measure was based partly on a similar questionnaire devised by Lankes (24). The score was the number of perseverative answers given. The complete questionnaire is as follows:

QUESTIONNAIRE ON PERSEVERATION

Underscore Yes if it describes your behavior better than No. No if it describes your behavior better than Yes.	Underscore	
1. If you have been disappointed, do you get over it easily?	Yes	No
2. Are you inclined to worry about things?	Yes	No
3. Do you go to sleep promptly upon going to bed?	Yes	No
4. Are you in the habit of keeping a diary or a journal? ...	Yes	No
5. Are you much given to day-dreaming?	Yes	No
6. Can you cram easily and retain the material for a quiz?	Yes	No
7. Are your dreams usually about some recent experience?	Yes	No
8. When you have some important task ahead of you (an examination or speech, for example) does it come frequently to your mind in the days just preceding? ..	Yes	No
9. Do you notice often that a tune, or a line of poetry, or a phrase, keeps frequently returning to your mind without an effort or intention on your part?	Yes	No
10. When writing a paper or working out a problem, can you drop it easily, pick up some other task, and put your mind easily upon the new task?	Yes	No
11. When you turn back to a task after an interruption, can you get it back into mind readily—do the former thoughts come back easily?	Yes	No
12. Do frequent changes in the routine of life annoy you? ...	Yes	No
13. After you have lived in one room or place for some time and are obliged to leave it, do you have difficulty in settling down to work in new quarters?	Yes	No

14. When you have once started something, do you feel uneasy if it is left incomplete?	Yes	No
15. After seeing a very tragic play or movie, does the emotion linger with you for some time after?	Yes	No
16. When studying or reading are you easily disturbed by what goes on around you?	Yes	No
17. Does it annoy you to have many different tasks or duties to look after?	Yes	No
18. Do you live by a daily program of work and relaxation?	Yes	No
19. When a personal remark has been made about you in your hearing (good or bad), do you take it to heart and remember it for a long time?	Yes	No
20. When some one makes you very angry, do you say little and have the anger linger with you disturbingly for some time after?	Yes	No
21. When some one makes you very angry, do you "fly off in a rage" and give him a good panning then and there?	Yes	No

RESULTS OF THE ADMINISTRATION OF THE FOREGOING MEASURES OF PERSEVERATION

All of the perseveration measures except those marked with an asterisk were given to 78 students of the University of Oregon. These students were chiefly sophomores and juniors with ages ranging from 16 to 22 years. Fifty-six of this number were women and 22 were men. The measures marked with an asterisk were administered and devised by Dr. Conklin, and it is through his kindness that they are available for this study. They were given to only 56 of the 78 in the larger group so that a separate table of correlations will be presented for this group of 56. The scoring of all the measures was made to correspond to the theoretical concept of perseveration so that the signs (plus or minus) of the coefficients should be correct according to the interpretation given the characteristics of perseveration by former investigators. (Product-moment r was used in all calculations.)

THE INTERPRETATION OF THE RESULTS OF THIS INVESTIGATION IN THE LIGHT OF THOSE OF FORMER INVESTIGATORS

The results presented in Tables 1 and 2 would seem to be not in accordance with the general tendency of the results of preceding studies in perseveration if Spearman's interpretation of these results be accepted. Significant positive evidence of a broad general factor of perseveration certainly could not be drawn from the results of the present investigation. A few examples will, I think, serve to show that the evidence obtained by former investigators has been

TABLE 1

INTERCORRELATIONS OF PERSEVERATION MEASURES, $N=78$.(r_{24} and r_{35} were actually plus but became minus for perseveration)

	2	3	4	5	6	7	8	9	14	15
1. Dark-adaptation	.22	-.13	-.15	.08	.03	-.02	.02	.00	-.15	.21
2. Mean simple reaction	.35	-.50	-.13	.11	-.08	.19	.06	.22	-.04	
3. Mean choice reaction		.17	-.44	-.11	-.02	-.17	-.02	.06	-.11	
4. Sigma simple reaction			-.09	.06	-.11	-.08	-.12	.09	-.07	
5. Sigma choice reaction				.01	.03	-.14	-.17	.00	-.05	
6. Tapping rate						.01	-.17	.07	.17	-.11
7. Star-tracing							-.18	.03		-.15
8. Motor inhibition (<i>I</i>)								.45	.00	.13
9. Motor interference (<i>Sd-Id</i>)									.17	-.04
14. Narrative memory test (<i>IR</i>)										-.27
15. Narrative memory test (<i>IR-DR</i>)										
Average intercorrelation, $r = -.021$.										

TABLE 2

INTERCORRELATIONS OF PERSEVERATION MEASURES, $N=56^*$

	6	.8	9	10	11	12	13	14	16
1. Light-adaptation	— .07	.03	— .27	— .26	— .04	.13	— .19	— .27	— .07
6. Tapping rate		— .19	.08	.19	— .01	— .22	— .12	— .04	— .16
8. Motor inhibition (<i>I</i>)			.44	— .26	— .04	— .21	— .26	— .02	.11
9. Motor interference (<i>Sd—Id</i>)				— .04	.00	.19	— .01	— .20	.18
10. Changed order letters					.10	.11	.10	— .02	— .09
11. Cancellation						— .18	— .21	.03	— .25
12. Arithmetic shift, add to multiply							.39	— .05	.15
13. Arithmetic shift, subtract to divide								.17	.17
14. Immediate recall (<i>IR</i>)									— .08
16. Questionnaire									
Average intercorrelation, <i>r</i> = — .021.									

*The omission of some of the measures from this table was made necessary by the failure on the part of some of the subjects to complete all of the measures given.

greatly overestimated in its importance as an argument for the functional unity of perseveration.

Spearman points out that the work of Heymans and Brugmans (15) gives no evidence of the functional unity of perseveration (33, p. 295). Lankes gives 9 measures of perseveration to 45 cases. His measures are representative of perseveration in the sensory, motor, and "ideational" fields. Lankes' results, as interpreted by Spearman (33, p. 301), provide the answer to the "great question, namely, as to whether these mental activities are really inter-cor-

TABLE 3
INTERCORRELATIONS OF PERSEVERATION MEASURES, GROUP 2.
[70 subjects, from Bernstein (3)]

	1	2	3	4	5	6	7	8	9	10
1. Tap- ping	.751	— .368	.339	.088	— .065	— .096	— .053	.112	— .088	— .282
2. Inverted "S"	.882	— .119	.453	.220	.155	.278	.005	.112	.240	
3. Reverse stroke		.749	.309	.162	.023	.191	.226	.194	.078	
4. Mirror image			.933	.142	.111	.182	.282	.233	.198	
5. "it" test				.601	— .023	.051	.002	.008	— .008	
6. Alphabet (reverse order letters)					.712	.004	— .051	.208	.048	
7. Triangles						.702	.114	.143	.223	
8. Capitals							.621	.108	.130	
9. Horizontal vertical								.658	.142	
10. "ea" test									.364	
Averages	— .046	.108	.156	.222	.045	.042	.126	.103	.118	.076

related." This interpretation is based on the fact that most of the coefficients in Lankes' table are positive and none significantly negative; the average positive intercorrelation being .22 (45 cases). Spearman suggests that the low intercorrelations may be due to attenuation. On the other hand, the "high" intercorrelation (only one coefficient as high as .50) may be due to the influence of "g" or "speed" as was brought out by the work of Hargreaves (12).

Bernstein makes the statement in reference to the tests he selected for pooling that they "have an appreciable positive correlation with one another; the average correlation for Group 2 is .181, *P.E.* .081, and for Group 3 is .171, *P.E.* .086." These correlations may indicate some slight relationship between the measures intercorrelated but this relationship is by no means established with certainty. With *P.E.*'s of this size, it is quite conceivable that the real interrelationships are often zero, and, at any event, the percentage of common factors indicated by coefficients of this size is negligible especially when the reliabilities are fairly high.⁶ Bernstein's results are shown in Table 3.

An illustration of what may happen with correlation coefficients as low as those in the above tables from Bernstein is given by a comparison of the coefficient r_{12} in Group 3 (.302) and the same coefficient in Group 2 (— .368). This relationship (between tapping

⁶About 3% common factors (r^2) is indicated between two measures which correlate .18. For another interpretation see (28).

TABLE 4
INTERCORRELATIONS OF PERSEVERATION MEASURES, GROUP 3.
[60 subjects, from Bernstein (3)]

	1	2	3	4	5	6	7	8	9	10
1. Tap- ping .814		.302	.231	.264	.181	.062	.301	.214	— .104	— .208
2. Inverted "S" .846			.254	.221	.246	.148	.169	— .051	— .053	— .225
3. Reverse stroke .773				.269	.153	.091	— .092	.212	.116	— .141
4. Mirror image					.868	.301	— .170	.257	.290	.198
5. "it" test						.687	.061	.126	.146	.260
6. Alphabet (reverse order letters)							.775	.051	— .010	— .003
7. Triangles								.814	.202	— .115
8. Capitals									.674	.226
9. Horizontal vertical										.631
10. "ea" test										
Averages	.138	.112	.121	.170	.169	.024	.098	.134	.070	— .065

and inverted "S" measures) might be considered as "significantly" positive in one group, but in that case it becomes "significantly" negative in the other group. The actual significance of nearly all of the coefficients in the above tables (3 and 4) is brought into question by this observation, for only one coefficient is above this value which appears positive in one group and negative in the other.

Spearman makes a table of averages of the above tables from Bernstein and calculates the tetrad differences for the averaged table (33, p. 304). He finds the median value of said differences to be .014, while the theoretical probable error of sampling is .013. He concludes that "This excellent agreement indicates that, in general, the whole of the correlations derive from only one factor." This might be a valid conclusion, but it has no significance if the actual intercorrelations are so low as to show a practically negligible relationship.

Bernstein has two series of tests of "speed," an "X" series and a "Y" series. The "mixed" arrangement of the tests in the "X" series called for rapid changes in adjustment on the part of the subjects and thus tended to penalize those of the perseverating type. The orderly arrangement of the tests in the "Y" series was more favorable to the perseverator as no such rapid changes were involved. Hence "Y—X" was taken as a measure of perseveration. His results show that "Y—X" has an "extremely small correlation with slowness; this would tend to confirm the conclusions previously ar-

rived at that perseverance is not a main factor in producing slowness." From these same results, however, it will be seen that "Y—X" has no significant relationship with the other measures of perseverance used (average $r=.07$). Bernstein calls "Y—X" a measure of "perseveration" and draws conclusions regarding its relationship to "slowness" and also calls his other measures of "perseveration" valid measures, drawing conclusions from the latter results even when there is no evidence of relationship between "Y—X" and alleged measures of "perseveration." He takes for granted the unity of the factor of "perseveration" in the face of his own evidence which seems to deny it.

Spearman cites the following table (Table 5) from the work of Wynn Jones (no reference given) which is the most encouraging evidence of intercorrelation between perseverance available.⁷

TABLE 5
INTERCORRELATIONS OF PERSEVERATION MEASURES FROM WYNN JONES
(77 children about 12 years of age)

	1	2	3	4
1. Mirror-wise "S"		.455	.340	.560
2. Back-stroke digits	.455		.520	.515
3. Mirror-drawing	.340	.520		.465
4. "it" test	.560	.515	.465	

These results show the presence of some group factor or factors in the measures used. One cannot be certain that the group factor is that of perseverance even though some precautions were taken to determine the effect of motor dexterity and speed. [See Spearman (33, p. 297) and Kelley (22, p. 20).] "Speed" and "g" have been shown to cause an appreciable lowering of intercorrelation coefficients of perseverance measures by Hargreaves. These factors were not eliminated in the work of Jones, even though Spearman shows that "g" is not a large factor in these measures.

Hargreaves has employed some of the same measures as those used by Jones, but he does not obtain the same results with 139 cases. Some of his results are shown in Table 6.

⁷In his most recent report (1929) Wynn Jones (21) concludes that "Although evidence has been produced that mental inertia as a scientific concept is a factor operative in many processes, further research is needed to determine if and how far it is operative in others."

TABLE 6
INTERCORRELATIONS OF PERSEVERATION MEASURES INCLUDING "g"
AND A "SPEED" FACTOR
[139 cases from Hargreaves (12)]

	B	C	D	E	F
A. Towns test	.37	.26	.00	.04	— .13
B. Word-building		.26	.03	.00	.09
C. Tapping			.03	— .10	— .04
D. Reverse stroke				.13	.13
E. Inverted "S"					.30
F. "it" test					

The tetrad differences calculated from the above table result in 21% being over the $P.E. \times 2$, while only 8.8% are to be expected from the sampling error. When the "speed" factor (speed of copying prose and figures) and "g" are eliminated the above table becomes as shown in Table 7.

TABLE 7
INTERCORRELATIONS OF PERSEVERATION WITH "g" AND "SPEED"
PARTIALED OUT
[139 cases from Hargreaves (12)]

	B	C	D	E	F
A. Towns test	.16	.19	.04	.06	— .14
B. Word-building		.23	.05	.01	.09
C. Tapping			.02	— .08	— .01
D. Reverse stroke				.12	.12
E. Inverted "S"					.29
F. "it" test					

The tetrad differences calculated from Table 7 are all positive and 61% are over the $P.E. \times 2$. Hargreaves concludes as follows:

"We have to conclude that, as far as any evidence we can procure shows, no general perseveration factor is found in our tests; but that the perseveration tests fall into two distinct groups (A.B.C. and D.E.F.). The latter (D.E.F.) was used by Bernstein and was found by him to correlate with estimates of perseveration made by teachers."

When all the available experimental work on perseverative phenomena is given critical examination, we are forced to the conclusion arrived at by Kelley (22, p. 21) that the "data are hardly serviceable in proving the existence of a perseverative factor." A

"functional unity" *cannot*, therefore, be ascribed to all of the diverse behavior processes involved in the list of "measures of perseveration" presented at the beginning of this paper. The gap between the neurological process involved in the adaptation of the eye, e.g., and the neurological process involved in tunes "running in the head" can be bridged only by the processes of logic as far as experimental evidence is concerned.⁸

There is some significant positive evidence for the existence of a narrow group factor of motor perseveration. Wynn Jones' positive results were obtained from measures in the motor field and it was these measures that Bernstein used in obtaining significant positive relationships with teachers' ratings (these being a pooled r of .49 and .52). Considering the known unreliability of judges' estimates, these coefficients are of considerable significance. The evidence for a grouping of motor measures of perseveration given by Hargreaves was mentioned above. The correlation between the two measures of motor perseveration developed in this study [motor inhibition (I) and motor interference ($Sd-Id$)], holding discrimination ability constant, becomes .53 ($P.E.=.06$). This coefficient is reduced to approximately .43 when the overlapping factors are eliminated. It should be raised somewhat if it was possible to partial out the effect of the perseverator's disturbance by a new situation which would tend to cancel the intended score for perseveration (mentioned in the description of these measures). This relationship is about the same as that found between two intellectual functions such as arithmetic and opposites. The evidence for the definite establishment of a narrow group factor of motor perseveration is as yet only suggestive and will await further experimentation in order to be at all conclusive. [See especially (21).]

Measures must be developed which are sufficiently specific for perseveration. The measures used may be fairly reliable but not valid measures of perseveration. Some of the measures used in this study were scored so that other factors could easily have masked a considerable portion of the "perseverative tendency." Spearman points out (33, p. 302) that the correlation between perseveration and mania-melancholia might easily be zero due to the influence of

⁸Physiology presents only one phenomena, other than sensory adaptation, which could be in any way similar to perseveration (as far as the author is aware), and that is Wundt's and Ritter's tetanus occurring in some isolated structures at the make or break shock of a galvanic current.

"self control" being greater with "normals" and less for both manics and melancholics. If there is a "refractory phase" in the memory process as is suggested by Dodge (7) this would tend to cancel any perseverative tendency of this process. Furthermore there are several phenomena, especially in the memory process, which appear to be due to "perseveration" but which, upon careful analysis, may be due to an entirely different process [as was pointed out by Foster (9)]. Hence the most conservative conclusion to be drawn is that the establishment of perseveration as a broad group factor awaits the development of measures which will analyze and measure the function of perseveration itself, stripped clean of all the foreign aspects of the disparate functions in which the perseverative tendency alone is to be measured.

It is very interesting to note that the recent report of Wynn Jones work⁹ (21) leads to the same conclusions (in general) as does the evidence brought forward in this paper. It will be remembered that reference was made to the far-reaching conclusions of Spearman in regard to the relation of perseveration to introversion and that these conclusions, as well as those in regard to perseveration and melancholia, were very much in need of experimental support. A preliminary attempt to investigate experimentally these latter conclusions of Spearman was included in this study. Some of these results will also be found in accord with the findings of Wynn Jones.

THE RELATION OF PERSEVERATION TO INTROVERSION AND DEPRESSION

Wynn Jones recent report (21) of the measurement of perseveration in mania and melancholia, in which he fails to corroborate the earlier work of Wiersma, has been mentioned. Jones concludes that "It is further suggestive that the view that cases of mania may be regarded as perseverators is untenable."

A questionnaire measure of degrees of depression was developed for the measurement of this trait in "normal" individuals.¹⁰ The

⁹The writer was able to obtain this report only a few days before this paper was presented before the Psychology Group of the Western Division of the American Association for the Advancement of Science in the spring of 1929.

¹⁰A report of the formulation and results of the administration of the DE test is to appear in the *Journal of Abnormal and Social Psychology* in the near future. Some of the results have appeared in the *American Journal of Sociology* (16).

Brown-Spearman reliability of this measure proved to be .78 ($P.E.=.03$). The raw coefficient of validity based on the ratings of six selected judges pooled was .71 ($P.E.=.06$). This coefficient was raised to .95 when corrected for attenuation.

Degrees of introversion were measured by the use of Conklin's Extraversion-Introversion Interest Test (5). This measure was used because of its high reliability (.95) and because it is in harmony with the present author's conception of the nature of extraversion-introversion, namely, that it is the characteristic direction of attention or interest with reference to the subjective and objective environments.

The relations found between the measurement of introversion and depression by Conklin's Extraversion-Introversion Test (*EI*) and the author's Depression-Elation Test (*DE*) and measures of perseverance are summarized in Table 8.

The only coefficient of possible significance (being 4 times the $P.E.$) in Table 8 is that between slowness of tapping rate and depression as measured by the *DE* test. This relationship is thrown into a rather peculiar light when the coefficient between tapping rate

TABLE 8

CORRELATIONS BETWEEN MEASURES OF PERSEVERATION, INTROVERSION (*EI*), DEPRESSION (*DE*), AND INTELLIGENCE ("*g*")*
(60 cases)

Perseveration measures	<i>DE</i>	<i>P.E.</i>	<i>EI</i>	<i>P.E.</i>	" <i>g</i> "	<i>P.E.</i>
1. Light-adaptation	.14	.08	— .15	.08	.03	.08
2. Mean simple-reaction-time	.16	.08	.14	.08		
3. Mean choice-reaction-time	.06	.08	.07	.08	incomplete	
4. Sigma simple-reaction-time	.20	.08	.06	.08	record	
5. Sigma choice-reaction-time	.10	.08	.05	.08	of " <i>g</i> "	
6. Tapping rate	.27	.07	.16	.08	.26	.09
7. Star-tracing	— .10	.08	— .10	.08		
8. Motor inhibition (<i>I</i>)	— .20	.08	— .08	.08	.02	.08
9. Motor interference (<i>Sd—Id</i>)	.03	.08	.14	.08	— .04	.08
10. Changed order letters	— .10	.10	.12	.10	.19	.09
11. Cancellation	.09	.10	— .22	.09	— .09	.10
12. Arithmetic shift, add to multiply	.14	.10	.00	.10	— .05	.10
14. Narrative (<i>IR</i>)	.06	.08	.07	.08	.10	.08
15. Narrative (<i>IR—DR</i>)	.17	.08	.17	.08		
16. Questionnaire	— .06	.10	.08	.10	.01	.10
Averages	.06	.08	.03	.08	.09	.09

*The American Council on Education's "Psychological Examination," 1927 edition, was used as the measure of "*g*."

and "g" is taken into consideration, since the *DE* test shows a correlation of $-.15$ ($P.E.=.06$) with the same measure of "g." These results will bear further experimentation.

A raw coefficient of $.37$ ($P.E.=.07$) was obtained between depression and introversion as measured by the *DE* test and the *EI* test, respectively. This coefficient is raised to $.43$ when corrected for attenuation. (These results are discussed elsewhere.)¹¹ This is indicative of some relationship between depression in characteristic mood and attitude and introversion in characteristic direction of attention or interest. This relationship might be stated in terms of about 18% (r^2) common factors in the measures employed. [For another interpretation see (28).]

The remainder of the coefficients in Table 8 are of negligible magnitude. These results furnish no evidence for a positive relationship between perseveration and introversion or between perseveration and depression.

GENERAL CONCLUSIONS

1. The results of the attempts to measure the perseverative tendency in various kinds of behavior processes taken from this study as well as from the results of other investigators fails to support the hypothesis of a broad group factor of "perseveration" or "mental inertia."

2. There is some evidence of significance for the existence of a narrow group factor of motor perseveration participating in a number of disparate motor processes which require a more or less rapid shift from one "pattern" of response to another "pattern" within the same general type of response (e.g., the "it" test and the changed-order letters measure).

3. The relationship between measures of perseveration and a measure of depression was investigated with no significant positive results.

4. The relationship between measures of perseveration and a measure of introversion was investigated also with no significant positive results.

5. It is necessary that measures of perseveration be developed which will be sufficiently specific for perseveration to eliminate the masking of the perseverative tendency by the many other factors

¹¹See the report to appear in the *Journal of Abnormal and Social Psychology* in the near future.

which might influence the score before any absolutely definite conclusions can be arrived at as to the nature or existence of the hypothetical functional unity "perseveration."

REFERENCES

1. BAIR, J. H. The practice curve. *Columbia Univ. Contrib. Phil., Psychol.*, etc., 1902, **9**, No. 3.
2. BERGSTROM, J. A. Experiments on physiological memory by means of the interference of associations. *Amer. J. Psychol.*, 1893, **5**, 356-369.
3. BERNSTEIN, E. Quickness and intelligence. *Brit. J. Psychol., Monog. Suppl.*, 1924, **3**, No. 7. Pp. vi+55.
4. BILLS, A. G. On the lag or "inertia" in the rise of the visual sensation. *Psychol. Monog.*, 1920, **28**, No. 127. Pp. 101.
5. CONKLIN, E. S. The determination of normal extravert-introvert interest differences. *Ped. Sem.*, 1927, **34**, 28-37.
6. CULLER, A. J. Interference and adaptability. *Arch. Psychol.*, 1912, **3**, No. 24. Pp. v+80.
7. DODGE, R. A note on Professor Thorndike's experiment. *Psychol. Rev.*, 1927, **34**, 237-240.
8. DODD, S. C. The theory of factors. *Psychol. Rev.* 1928, **35**, 211-233.
9. FOSTER, W. S. On the perseverative tendency. *Amer. J. Psychol.*, 1914, **25**, 393-428.
10. GRAY, J. Perseveration as a test of the quality of intelligence. *Rep. Brit. Asso. Adv. Sci.*, 1910, **80**, 805-806.
11. GROSS, O. Die cerebralen Sekundärfunktionen. Leipzig, 1902. Pp. 69.
12. HANSEN, C. F. Serial action as a basic measure of motor ability. *Psychol. Monog.*, 1922, **31**, No. 40, 320-382.
13. HARGREAVES, H. L. The "faculty" of imagination. *Brit. J. Psychol., Monog. Suppl.*, 1927, No. 10. Pp. 74.
14. HECHT, S. The nature of foveal dark adaptation. *J. Gen. Physiol.*, 1919-1920, **2**, 499-517; 1921-1922, **4**, 330-340.
15. HEYMANS, G., & BRUGMANS, H. Intelligenzprüfungen mit Studierenden. *Zsch. f. angew. Psychol.*, 1913, **7**, 317-331.
16. JASPER, H. H. Optimism and pessimism in college environments. *Amer. J. Sociol.*, 1929, **34**, 856-873.
17. JERSILD, S. T. Mental set and shift. *Arch. Psychol.*, 1927, **14**, No. 89. Pp. 81.
18. JOHNSON, H. M. Reaction-time measurements. *Psychol. Bull.*, 1923, **20**, 562-589.
19. JONES, L. W. Perseveration. *Rep. Brit. Asso. Adv. Sci.*, 1915, 698.
20. ———. An investigation into the significance of perseveration. *J. Ment. Sci.*, 1928, **74**, 653-659.
21. ———. Individual differences in mental inertia. *J. Indus. Psychol.*, 1929, **4**, 282-294.
22. KELLEY, T. L. Crossroads in the mind of man. Stanford University, Calif.: Stanford Univ. Press, 1928. Pp. 245.
23. KNIGHT, F. B., & FRANZEN, R. B. Pitfalls in rating schemes. *J. Educ. Psychol.*, 1922, **13**, 204-213.

24. LANKES, W. Perseveration. *Brit. J. Psychol.*, 1915, 7, 387-419.
25. McCOMAS, H. C. Apparatus for recording of continuous discrimination reactions. *J. Exper. Psychol.*, 1917, 17, 171-177.
26. McDONOUGH, M. R. The empirical study of character. Part II. *Stud. Psychol. & Psychiat.*, 1929, 2, 147-213.
27. MÜLLER, G. E., & PILZECKER, A. Experimentelle Beiträge zur Lehre vom Gedächtniss. *Zsch. f. Psychol.*, 1900, Erg. 1. Pp. xiv+300.
28. NYGAARD, P. H. A percentage equivalent of correlation. *J. Educ. Psychol.*, 1926, 17, 86-92.
29. REXROAD, C. N. A continuous multiple choice reaction apparatus. *J. Exper. Psychol.*, 1925, 8, 325-336.
30. SEASHORE, R. H. The Stanford motor skills unit. *Psychol. Monog.*, 1928, 39, No. 178, 51-66.
31. SHEN, E. The validity of self estimates. *J. Educ. Psychol.*, 1925, 16, 104-108.
32. SNODDY, G. S. Learning and stability. *J. Appl. Psychol.*, 1926, 10, 1-36.
33. SPEARMAN, C. The abilities of man. London, New York: Macmillan, 1927. Pp. xxiii+415.
34. ———. The uniqueness of "g." *J. Educ. Psychol.*, 1929, 20, 212-216.
35. WOODROW, H. The measurement of attention. *Psychol. Monog.*, 1914, 17, No. 76. Pp. 158.
36. WIERSMA, E. Die Sekundärfunktion bei Psychosen. *J. f. Psychol. u. Neur.*, 1906, 8, 1-24.

University of Oregon
Eugene, Oregon

LA PERSÉVÉRATION EST-ELLE UNE UNITÉ FONCTIONNELLE PARTICIPANT À TOUS LES PROCESSUS DU COMPORTEMENT?

(Résumé)

On a donné seize mesures de "persévération" à 78 étudiants universitaires. L'intercorrélation moyenne entre ces mesures a été pratiquement de nulle signifiante. Ces résultats, avec l'examen critique des autres investigations dans ce domaine, donne la conclusion à l'auteur qu'il n'y a pas encore de l'évidence satisfaisante de la présence d'un grand facteur collectif de "persévération" ou "inertie mentale." Ceci ne veut pas dire que la "persévération" comme grand facteur collectif a été réfutée, car il est tout à fait possible que les mesures employées soient entièrement insuffisantes pour la mesure de la persévération elle-même, libre des éléments étrangers qui tendent à annuler le résultat pour la persévération.

Il se montre quelque évidence dans cette étude de la présence d'un petit facteur collectif de persévération motrice qui corrobore les résultats de Hargreaves et de Bernstein à cet égard.

On a développé une mesure de dépression-exaltation à employer avec des sujets normaux laquelle a montré un coefficient corrigé de validité de 0,95, P.E. 0,03, et un coefficient de constance Brown-Spearman (pair et impair)

de 0,78, P.E. 0,08. La corrélation moyenne entre l'introversion, mesurée par le Test D'Intérêt Conklin de l'Extroversion-Introversion, et ces mesures de persévération a été de 0,03, P.E. 0,08. Ces résultats ne soutiennent pas du tout l'hypothèse que la persévération a une relation avec la dépression et que la persévération peut être identifiée avec l'introversion.

JASPER

IST DAS VERHARREN EINE FUNKTIONELLE EINHEIT ALLER PROZESSE DES VERHALTENS

(Referat)

Man gab 78 Collegestudenten 16 verschiedene Messungen des "Verharrens" (perseveration). Die durchschnittliche Wechselbeziehung zwischen diesen Massen war so gering, dass die Berechnung der tetraden Differenzen fast bedeutungslos wird. Diese Ergebnisse zusammen mit einer kritischen Prüfung anderer Untersuchungen dieses Gebietes, führen den Verfasser zur Schlussfolgerung, dass es bis jetzt noch keine befriedigende Beweise für das Bestehen eines umfassenden Gruppenfaktors, "des Verharrens," oder der "geistigen Trägheit" ("mental inertia") gibt. Das bedeutet aber nicht, dass das Fehlen des "Verharrens" als umfassender Gruppenfaktor bewiesen wurde; denn es ist ganz wohl möglich, dass die angewandten Masse ganz unzulänglich sind für die Messung des Verharrens überhaupt, abgesehen von fremden Elementen welche das Mass des Verharrens aufzuheben trachten.

In der vorliegenden Arbeit liegt etwas Beweismaterial vor für die Existenz eines engern Gruppenfaktors motorischen Verharrens, welches die Ergebnisse von Hargreaves und Bernstein in dieser Hinsicht bestätigen.

Es wurde ein Mass für depressive und frohe Erregung für normale Personen entwickelt, welches einen korrigierten Gültigkeitskoeffizienten von 0,95, mit einer Fehlerquelle (P.E.) von 0,03, und einen Brown-Spearman Zuverlässigkeitskoeffizienten (gerade gegen ungerade) von 0,78 mit Fehlerquelle von 0,08 aufweist. Die durchschnittliche Korrelation zwischen Introversion, gemessen mit dem Conklin Extraversion-Introversion-Interessen-Test, und diesen Massen des Verharrens war 0,03 mit Fehlerquelle von 0,08. Diese Ergebnisse unterstützen die Hypothese in keiner Weise, dass das Verharren mit Depression in Beziehung stehe, und dass das Verharren mit Introversion identifiziert werden könne.

JASPER

A STUDY OF BIRTH ORDER AND BEHAVIOR*

IRA S. WILE AND ELINOR NOETZEL

The family constellation has been made an important factor of child development by Alfred Adler and his followers. Their theory practically creates a determinism of the characteristics based upon order of birth. It is important to ascertain if the theoretic assumptions of influences of ordinal position are borne out by experience. For this reason we have undertaken a study to inquire into any possible relationship between the overt problems of children and family constellation. It is obvious that if order of birth is a dominant element in personal and social evolution there should be marked reflections of it in child behavior. In many instances the accentuation of conditions born of ordinal position should lead to pronounced behavior disorders. For these reasons we have sought to consider statistically more or less definite types of children in relation to their order of birth.

To check our study of the children we have also considered a group of adults whom we have classified somewhat similarly to the children on the basis of behavior reactions. We desired to note what similarities might exist relating to ordinal position in the family. We recognize, however, the marked differences in the selection of the two groups as will be indicated later.

Five hundred children were selected in consecutive order from the records of the Children's Health Class of Mount Sinai Hospital. To eliminate any problems that might be related to mental deficiency, all children with IQ's lower than 70 were eliminated. We did this despite Dayton's study showing that "The order of birth of the retarded or mentally defective child apparently is of little significance." We further rejected children in the borderline group whose mental levels were much below the familial level because we felt that these children were facing the same social problems as those of the obviously mental deficient. Finally, we discarded the histories of all children suffering from organic brain defects or diseases.

*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office February 25, 1930.

For the purpose of consideration we adopted a somewhat arbitrary classification in that we have grouped all our children according to explosive or withdrawn types, enuresis, and delinquency.

Tabulations were made of all the children in terms of oldest, middle, and youngest child, only child, and twins. We have considered all oldest children as only children when there was a difference of more than four years between him and the second child, for socially he has been *de facto* an only child during the important formative years of life as stressed by Adler. We further considered the children as living rather than as children born unless the death was comparatively recent and the deceased had played a definite part in the family constellation. Patently a stillborn child as the first child should be rejected in this consideration of the problems of the second-born child who would be, if living, the oldest child of a family, or, if there were no other child, the only child. No doubt the attitude of the parents toward the stillbirth is not negligible, but the stillborn can exert no pressure from position.

In differentiating between the explosive and withdrawn types of personality we endeavored to study the personality and answer the question, "Did this child fight against reality or did he withdraw from it?" Was the child centrifugal in relation to his environment, attacking it and meeting it aggressively, or was his tendency to have centripetal activity with retreat from his environment? Hence we have included in the explosive type of personality children with such reactions as fighting, bullying, tantrums, so-called disobedience, hyperactivity, cruelty, and unpleasant social aggressiveness leading to unpopularity. We have included under the withdrawn type those presenting evidence of seclusiveness, shyness, day-dreaming, marked fears, hypochondriasis, psychic constipation, vomiting, headaches, and the like, dependence on mother, speech defects not due to faulty training, certain forms of psychogenetic chorea, selecting only those which recovered under psychotherapy, and tics when there was no indication to suggest an explosive type of personality.

We recognized that masturbation ordinarily would be placed as an activity representative of the withdrawn group. Most of the children treated for this habit are thus segregated. Although we recognize that this may be open to objection, we feel that until a term is coined which permits differentiation between childish curiosity and real neurotic practice, masturbation should not be regarded as typical of any personality group.

Enuresis was selected as a general heading because it represents a definite habit which is easily recognized and could be tabulated. We appreciate that here there may be criticism because of the various questions which still arise as to whether enuresis is an evidence of neuropathic constitution or is due to faulty training and inadequacy of habit formation. The same criticism may be applied to one of the groups of adults later to be considered.

The connotation of the word delinquency presented a real difficulty. To say delinquent did not mean getting caught. In our classification we tried to limit our judgment to a consideration of those acts which would have brought the children ordinarily into the Children's Court if they had been apprehended. On studying our case records we found, except in three cases where the children were connected with stealing, our truants could be classed with either of our reaction groups, that is, some children were truants because they found the reality of school life too hard and withdrew from it, while others truanted as part of a rebellion against authority. In the delinquent group we included stealing and sexual activity, when the act involved another person and endangered the community. There were thus in the group two girls illegitimately pregnant, with records of promiscuity, and one child who persistently taught masturbation to another one. Destructiveness when it included the destruction of property of others was included. There was one instance of a child who practiced mahem on his school fellows.

The age of the the groups merits consideration. Our reaction groups included children from the age of 2 to 16 years. We limited our enuretic group to children between the ages of 4 and 16 so as to eliminate any possible hangovers of infantile incontinence. The age distribution of the delinquent group was from 9 to 16 years, as such behavior is more common in older groups. We are thoroughly aware that we are open to criticism because our groups are not of the same age distribution, but the distribution was not made arbitrarily, but naturally fell into these age groups for the reasons mentioned.

The cultural and racial background of our entire group merits consideration. The children came from all five boroughs of New York City. They were of higher economic and social status than is common to the average clinic patient. Our ratio between our explosive and withdrawn types of 100 to 74 speaks well for the

parental insight and the degree of mental hygiene education which they reflect. Approximately 60% of the children had IQ's of 100 or above. Seventy-two per cent of the children were of Hebrew parentage.

It appeared desirable to contrast the traits of the immature subjects with some similar traits appearing among adults. We are deeply indebted and here acknowledge our gratefulness to Dr. Franklin G. Ebaugh, Director of the Colorado Psychopathic Hospital, for the use of his exceptionally detailed records which gave a very clear picture of the patient's childhood personality.

We made use of 355 case records of hospitalized adults, functionally neurotic, psychotic, and delinquent patients. We endeavored again to eliminate all mental defectives, not only by excluding all patients who had a diagnosis of psychosis with mental deficiency, but also those whose earlier school records led us to suspect a definite difficulty in the intellectual sphere.

We classified the adult group according to the same method with reference to family constellation as we did in the childrens' group. We further established four similar categories of explosive or withdrawn types, delinquents, and an addict group. The last-named we substituted for the enuresis group on the basis of a possible functional similarity in that it related to a habit partly within personal control.

In the explosive group we included all manic depressives in the manic phase and all accessible depressives, all psychoneurotics and psychopathic personalities whose illness expressed itself in the form of tantrums. In the withdrawn or schizoid group we gathered the patients with dementia praecox, inaccessible depressives, anxiety neurotics, psychoneurotics with hypochondriacal trends, and reactive depressives.

In the addict group we included those suffering from narcotism and alcoholism, excluding however, all alcoholics who had been diagnosed as constitutional psychopathic inferiors. It was our desire to consider this group in the light of a habit problem, and we did not wish to enter into any discussion concerning the relation of alcoholism to neuropathic constitution.

In the delinquent group we have included constitutional psychopathic inferiors and other persons accused of crime. Perhaps it is desirable, inasmuch as we are considering the records of Dr. Ebaugh, to understand his meaning of the term constitutional psychopathic

inferior. He defines such a person as an individual who does not profit by experience, through some defect which is not intellectual. This definition is partially qualified by stating that early habit training could prevent a number of such persons from becoming social problems. According to the present Colorado law, a person accused of crime who plans to present an insanity defense may be observed at the State Psychopathic Hospital for 30 days. Hence, in our delinquent group there are persons accused of murder, embezzlement, grand larceny, forgery, bigamy, sex offenses, malingering with intent to defraud, petty larceny, robbery, and vagrancy. In this group we have included no persons found to be suffering from a psychosis. When a diagnosis of constitutional psychopathic inferiority was made in connection with alcoholism, we included them in this group as they always evidence anti-social traits. A few of the individuals who had committed major crimes were diagnosed as normal, and we feel justified in including them as they present the same anti-social problems that our delinquent children presented.

The age distribution of our adult group varied from 16 to 50 years.

The cultural and racial background of the patients is diverse because they came from all over the Middle West. They were, however, as in the children's group, of higher social and economic status than the average clinic or public hospital patient. Eighty per cent of the patients were of American stock and of Nordic descent. The median number of children born to the families of the children's groups was 3.6, and of the adult group 4. It is noteworthy that of the children the total possible number of patients in all the families was 1827, of whom 27.9% were seen; among the families of the adult group there were 1503 children, of whom 23.6% were in the Colorado institution. Making allowance for the difference in the age groups and the possibilities of death and the like, the distinction of percentages of the two groups is exceedingly marked and suggests perhaps that our mode of selection is not entirely satisfactory.

Let us first consider the children's group (Table 1).

It is to be noted that there is a marked similarity between the percentage of oldest children who belong to the explosive and withdrawn types and this obtains also for the youngest and the middle children. There is perhaps a slightly increased number of only children who appear to be explosive. Among the enuretics there was a larger percentage of the oldest children, but there was not a

TABLE 1
CHILDREN'S GROUP

	Explosive		Withdrawn		Enuretic		Delinquent		Total	
	N	%	N	%	N	%	N	%	N	%
Oldest	76	34.0	55	33.3	38	42.0	16	66.0	185	37.0
Middle	60	27.0	52	31.6	24	26.6	4	16.0	140	28.0
Youngest	56	25.0	43	26.0	19	21.0	3	12.0	121	24.0
Only	26	11.7	11	6.6	9	10.0	1	4.0	47	9.0
Twins	4		2		0		0		7	1+
	222	44.4	164	32.8	90	18.0	24	4.8	500	99+

marked number of only children. The delinquent children showed a larger proportion of oldest children, mainly because delinquency is more common among the pre-adolescents and adolescents, and our group age limit was 16 years. Hence there would be a natural gravitation of the oldest children in this group. It is apparent, however, that with the exception of the delinquent group, the proportion of oldest, youngest, middle, and only children in each class approached rather closely the proportion of these groups in the total number of children.

One notices that the total number of children of the explosive type constituted 44.4% of the total, and one would expect more of these children as they are more difficult for parents to manage and represent difficulties with children at almost any age. One would also expect a high percentage of oldest children to be brought in with enuresis as the condition is more disturbing to the family as the children grow older, particularly if there be also children under four years who are still wetting. By the exclusion of those under four years, we eliminated a large number of the youngest and some middle children from our tabulation.

An effort was made to compare this special group of children with the ordinary admissions to the pediatric clinics. We appreciate the cooperation of Dr. John Levy who gathered the data for us concerning a group of unselected children who were coming for the treatment of physical ailments. Unfortunately the major collection of facts was during the summer months rather than distributed over the entire year. Hence there is some invalidity of the statistics owing to the seasonal variation of diseases among children. There would tend to be more illness among youngest children than oldest children because of the prevalence of digestive difficulties during the

TABLE 2
ADMISSIONS TO PEDIATRIC CLINICS

	Males		Females		Total	
	N	%	N	%	N	%
Oldest	33	23.0	23	16.9	56	19.0
Middle	46	32.0	43	31.0	89	30.0
Youngest	44	33.0	57	41.0	101	39.0
Only	20	14.0	16	11.0	36	12.0

heated term. We present the tabulation, however, as it stands (Table 2).

Before considering the relations of the admission group to the children's group perhaps it is advisable to present the tabulation for the adult group (Table 3).

TABLE 3
ADULT GROUP

	Manic		Schizoid		Addict		Delinquent		Total	
	N	%	N	%	N	%	N	%	N	%
Oldest	29	24	43	24	8	27	9	34	89	25
Middle	41	33	59	33	3	10	9	34	112	32
Youngest	29	25	42	23	7	24	4	15	82	24
Only	20	16	32	18	11	39	4	15	67	18
Twins	2		3		0		0		5	1
	—	—	—	—	—	—	—	—	—	—
	121	34	179	49	29	8	26	7.4	355	100

As might be expected, in contradistinction to the juvenile group, there was a larger percentage of the withdrawn personalities under hospital care. One recognizes the reason for the difference in the total 44.4% of explosive children in the group as compared with 34% of the manics and the 32.8% of the withdrawn children as contrasted with 49% of the schizoid. The differences between the personalities of enuretics and addicts is obvious from the very nature of the distinction of the two types of complaints of the age groups involved. The difference in the delinquent group is less marked but obviously represents the difference between a group who are young as compared with a group that is older and has had larger opportunities for anti-social conduct.

Let us now place in columns our three groups and consider the ordinal positions by percentages, omitting the twins.

TABLE 4

	Adult group %	Pediatric clinic %	Children's group %
Oldest	25	19	37
Middle	32	30	28
Youngest	24	39	24
Only	18	12	8

One notes at once some similarity of the percentages of middle children in the three groups, and it is also worthy of recognition that in the adult and juvenile group the percentage of youngest children is identical. While there is a variation in the number of oldest children in the adult group and juvenile group, it is worthy of notice that the oldest and only children together constitute 43% of the adult group and 45% of the children's group. The low percentage of only children in the children's group can be understood in part by the fact that only children would be there only when their behavior sorely taxed the family. The increased number of oldest children in the juvenile group is readily appreciated as they are more likely to be disturbing than younger children, thus placing a greater pressure upon the mother to secure relief, particularly from such conditions as enuresis and delinquency.

To simplify our tabulation, let us omit the only children and twins. This will reduce the total number of children in our group to 466 and the number of adults to 283.

The values of our group change slightly. Of the children there were 192 (43%) of the explosive type, 150 (33.6%) of the withdrawn type, 81 (18%) of the enuretic type, and 23 (5%) of the delinquent type, contrasted with 99 (35%) of the manic type, 144 (50.9%) of the schizoid type, 18 (6.4%) addicts, and 22 (7.8%) delinquents. These groups may be tabulated as shown in Table 5.

TABLE 5

		Oldest			
explosive	39.6	withdrawn	36.6	enuretic	47.0
manic	29.0	schizoid	30.0	addict	44.0
		Middle			
explosive	31.0	withdrawn	35.0	enuretic	30.0
manic	41.0	schizoid	41.0	addict	16.6
		Youngest			
explosive	30.0	withdrawn	28.6	enuretic	23.0
manic	29.0	schizoid	29.0	addict	39.0
				delinquent child	69.6
				delinquent adult	41.0
				delinquent child	17.4
				delinquent adult	41.0
				delinquent child	13.0
				delinquent adult	18.0

TABLE 6

	N %		N %		N %		Total N %							
Oldest	I manic	29	24.0	schizoid	43	24.0	addict	8	27.0	delinquent	9	34.0	89	25.0
	II explosive	76	34.0	withdrawn	55	33.3	enuretic	38	42.0	delinquent	16	66.0	185	37.0
Middle	I manic	41	33.0	schizoid	59	33.0	addict	3	10.0	delinquent	9	34.0	112	32.0
	II explosive	60	27.0	withdrawn	52	31.6	enuretic	24	26.6	delinquent	4	16.0	140	28.0
Youngest	I manic	29	24.0	schizoid	42	27.0	addict	7	24.0	delinquent	4	15.0	82	24.0
	II explosive	56	25.0	withdrawn	43	26.0	enuretic	19	21.0	delinquent	3	12.0	121	24.0
Only	I manic	20	16.0	schizoid	32	19.0	addict	11	39.0	delinquent	4	15.0	67	18.0
	II explosive	26	11.7	withdrawn	11	6.8	enuretic	9	10.0	delinquent	1	4.0	47	8.0
Twins	I manic	2		schizoid	3								5	1.0
	II explosive	4		withdrawn	3		enuretic	0		delinquent	0		7	1.0
Total	I manic	121	34.0	schizoid	179	49.0	addict	29	8.0	delinquent	26	7.4	355	100.0
	II explosive	222	44.4	withdrawn	164	32.8	enuretic	90	18.0	delinquent	24	4.8	500	100.0

This tabulation, perhaps, does not suffice to bring out variations sharply, but there are a few suggestive figures, as, for example, the similarity of percentages for the youngest child in the explosive and manic groups and in the withdrawn and schizoid types. If one may assume that all the patients came from families with three children, the operation of the ordinary law of chance should give an equal proportion to each of the three groups, the oldest, middle, and youngest. We shall treat this phase of the problem of chance in more detail later. Our numbers are not large, but one notes this tendency in the explosive and withdrawn groups and manic and schizoid groups. Because of the fact that many families have more than three children, there should be a tendency for a slightly higher proportion of middle children, and this is observable in the manic and schizoid groups. Inasmuch as only children and twins constitute 19% of our adult group and 10% of the children's group, it is perhaps unfair to disregard them, and Table 6 sets forth the correspondences.

Reducing Table 6 to bases of contracted groups, we have Tables 7*A*, *B*, *C*, and *D*; a slightly larger grouping is presented in Tables 8*A* and *B*.

TABLE 7

<i>A. Manic and Explosive</i>			<i>B. Schizoid and Withdrawn</i>		
	Manic %	Explosive %		Schizoid %	Withdrawn %
Oldest	24.0	35.0	Oldest	24.0	33.3
Middle	33.0	27.0	Middle	33.0	31.6
Youngest	25.0	25.0	Youngest	23.0	26.0
Only	16.0	11.7	Only	18.0	6.6
Oldest and only	40.0	45.7	Oldest and only	42.0	40.0

<i>C. Addict and Enuretic</i>			<i>D. Delinquency</i>		
	Addict %	Enuretic %		Adult %	Child %
Oldest	27.0	42.0	Oldest	34.0	66.0
Middle	10.0	26.6	Middle	34.0	16.0
Youngest	24.0	21.0	Youngest	15.0	12.0
Only	3.9	10.0	Only	15.0	4.0
Oldest and only	30.9	52.0	Oldest and only	49.0	70.0

TABLE 8

*A**

	Manic %	Schizoid %	Explosive %	Withdrawn %
Oldest	24.0	24.0	34.0	33.3
Middle	33.0	33.0	27.0	31.6
Youngest	25.0	23.0	25.0	26.0
Only	16.0	18.0	11.7	6.6
Oldest and only	40.0	42.0	45.7	40.0

B†

	Addict %	Delinquent %	Enuretic %	Delinquent %
Oldest	27.0	34.0	42.0	66.0
Middle	10.0	34.0	26.6	16.0
Youngest	24.0	15.0	21.0	12.0
Only	39.0	15.0	10.0	4.0
Oldest and only	66.0	49.0	52.0	70.0

*Note percentages of youngest and middle correspond very well; also oldest and only.

†Youngest correspond in all groups and deviation for oldest is not great.

Table 7*A*, contrasting the manic and the explosive types, shows an identity of percentage distribution of youngest children, with a larger percentage of only children in the manic group than in the explosive, and a larger number of oldest children in the explosive group. In 7*B* there is a fair degree of agreement with the youngest group and the middle group. If we add the oldest and only groups together, we have 42% in the schizoid group, as opposed to 40% in the withdrawn group. This compares fairly well with the 40% and 45.7% for the manic and explosive groups. In 7*C*, obviously, there are marked variations due to the difference in the nature of the groups, but even here one sees some similarity in the percentages of the youngest members of the group. In 7*D* the two delinquent groups not only contained few individuals but the very nature of the make-up of the groups brings about marked disparities.

Table 8*A* indicates a reasonably high correspondence for the youngest and middle of the juvenile and adult groups. The difference between the percentages for the oldest children lies rather in the increased number of only children in the manic and schizoid groups as compared with the explosive and withdrawn groups. In 8*B* the only ordinal group in which there is any degree of approach

is the correspondence in the youngest group. The deviation in the oldest group is fairly well marked for reasons that have already been referred to.

If, as was suggested above, the families of our two groups were all of the three-children type, then the distribution for each ordinal position would be 33 1/3%. If the families were of the four-children type 25% would be the groups, respectively, for the oldest and youngest, and 50% for the middle group.

In a group of children from families with the average number of children between 3 and 4, the percentages for the oldest and the youngest would be between 25% and 33 1/3%, while the middle group would vary between 33 1/3% and 50%.

The average number of children in our group of children at Mount Sinai Hospital was 3.6 per family, and the average number in Dr. Ebaugh's adult group was 4 per family. Table 9 shows the distribution of the numbers of children in the family.

The consideration of our adult group in Table 3 indicates that the youngest and oldest sections approximate the theoretic 25% that one would expect in 4-child families for all save the delinquent group. If one combines the middle and only children, one has 49% for the manic group and 51% for the schizoid group, 49% in the addict group, and 49% in the delinquent group. Considering the distortion that arises from the inclusion of families that are not strictly 4-children families, there appears to be little irregularity in

TABLE 9

	Children's group	Total children	Adult group	Total adults
1-child family	37	37	67	67
2-child family	129	258	55	110
3-child family	113	339	62	186
4-child family	90	360	31	124
5-child family	42	210	50	250
over-5-child family	89	623	—	—
	500	1827		
6-child family			30	180
7-child family			29	203
8-child family			17	136
9-child family			7	63
10 or more			7	84
			355	1403

the distribution. When we consider the juvenile group (Table 1), there is again an illustration of reasonable approach to the theoretic basis in that in both the explosive and withdrawn type the proportion is what would be expected in a 3-child family. The youngest group shows reasonable approach but is nearer the expected percentiles for 4-children families. This distortion can be accounted for partially by the fact that the average number of children was not exactly 3 but 3.6 per family. One would expect a somewhat lower percentage of the oldest children in families with 3.6 children.

If we consider our children's group, omitting the only children and twins (Table 5), we find that in our explosive group the oldest form 39.6%, the middle 31%, and the youngest 30%; in the withdrawn group the oldest form 36.6%, the middle 35%, and the youngest 28.6%. Here there is some approach then to the theoretic distribution of families with 3 children, with little more distortion than one would expect considering that the number of children per family was not actually 3. If we consider the adult group with the same proviso, there is a slight shifting of the figures. In the manic group the oldest would constitute the middle 41%, and the youngest 29%. The similarity in percentages for the oldest and the youngest suggests, of course, that death has had definite effects upon the usual distribution of these age classes.

It is evident, therefore, from a consideration of our two groups that the distribution of the oldest, middle, and youngest subjects approximates the theoretic distribution that one would expect with our juvenile group coming from families with 3.6 children and our adult group coming from 4-children families.

If we compare our groups in order of birth with other tabulations, some interesting facts evolve. In Cattell's study of *American Men of Science*, he indicates the percentages of children born in various ordinal positions in families of different numbers. Considering men of science coming from families with three children, he finds that 44% were first-born, 31.2% were second-born, and 24.8% were third-born. In families with four children the percentage was somewhat altered, 36.1% being first-born, 22.4% being second-born, 21.8% being third-born, and 19.7% being fourth-born. In two-children families 57.4% were first-born and 42.6% were second-born.

C. B. Willis, in a comparison of Stanford-Binet IQ's of 219 pairs of first- and second-born children, found a slight difference in favor

of the second-born. This, however, is not a universal finding. Terman, for example, in his study of children with high IQ's found some interesting facts which I shall tabulate for both sexes. In his group of 574 children characterized as being in the intellectual genius group, 88 (15.3%) were only children, 185 (32%) were oldest children, 143 (25%) were middle children, 158 (27.5%) were youngest children. It is worthy of mention that he notes his groups for families of only three children and four children and which were found respectively to consist of oldest 36.9 and 33.0%, middle 31.9% and 42.2%, and youngest 31.2% and 24.7%. These figures are somewhat in harmony with Cattell's distribution, but there is again indication of the reasonable similarity of the totals with our own figures, considering the high degree of special selection in the Terman group. The Cattell groups, like the Terman, represented individuals of exceedingly superior IQ's as contrasted with the normal or slightly superior groups in our juvenile and adult group. There is evident throughout a tendency for the youngest groups to be about the same.

Variations in figures are shown, for example, in the study of Breckenbridge and Abbott of 585 delinquent boys in Chicago, of whom 23.9% were oldest children, 60% middle children, and 12% were youngest children. Reynolds in the Massachusetts Habit Clinic for Problem Children found 27% oldest, approximately 32% middle, 20.8% youngest, and 11.8% of only children, with the additional figure of those presented as adopted children.

TABLE 10

	Census	Children	Adult	Terman			Cattell		Reynolds
				Three- children families	Four- children families	All families	Three- children families	Four- children families	
Oldest	23	37	25	36.9	33.0	32	44	36.1	27
Middle	26	28	32	31.9	42.2	25	31.2	44.2	42
Youngest	23	24	24	31.2	24.7	27.5	21.8	18.7	20.8
Only	28.5	9	18			15.3			11.8
Oldest and only	51.5	46	43			47.3			38.8

There are no data indicating the ordinal distribution of people in our population. In considering the order of birth per thousand mothers in 1922, one finds that 28.5% had the first birth, which would represent, of course, for them only children. Twenty-one and nine-tenths per cent had their second child; this does not neces-

sarily mean that the first child was living, so one cannot be certain that this represents the percentage of 2-children families. All the children born except only children, 71.5%, would have to be the youngest, but also theoretically 71.5% of the children would be the oldest providing there were no deaths. To have no middle children would mean that the family consisted of less than three children, and this total would represent 50.4%. Theoretically, then, the youngest group in families of more than three children would be 35.2% because the number of mothers with the third child born was 14.4%. This form of calculation, however, is inadequate. If we will take the following order of births per thousand mothers in 1922 according to the United States Census (Table 11), we may derive figures showing percentage distribution in the population of the oldest, middle, youngest, and only children as far as ordinal position is concerned, but without reference to the actual number of children living.

TABLE 11

Order of birth	1	2	3	4	5	6	7	8	9	10	11-27
	28.5	21.9	14.4	10.3	7.31	5.29	3.82	2.73	1.88	1.27	1.8

It is easy to compute that the families of the thousand mothers had in total approximately 3350 children born alive, or an average of 3.3 children per family. Calculations further reveal the distribution of the children born to be: oldest 23%, middle 26%, youngest 23%, only 28.5%.

If one considers an infant mortality rate of 70 per year, the total change would affect only 70 families of the one thousand and hence would not have a tremendous influence in distorting figures. It is significant that the average number of children ever born to the mothers of 1922 was 3.3. To be sure, one cannot be certain of the family group of three children or more, but there is a suggestion of the probable family grouping in the general population, even though the figures are not to be regarded as absolute.

If we now compare our groups (Table 12), we see that in the child group the middle and youngest groups are fairly in harmony with the theoretic distribution. There is a distortion in the figures for the oldest and only children because the census group considers the one year without relation to deaths, whereas our group represents life experience over many years. If, however, one adds the oldest

TABLE 12

	Census	Total Children	Total Adults	Explosive	Withdrawn	Manic	Schizoid
Oldest	23.0	37.0	25.0	34.0	33.3	24.0	24.0
Middle	26.0	28.0	24.0	27.0	31.6	33.0	33.0
Youngest	23.0	24.0	32.0	25.0	26.0	25.0	23.0
Only	28.5	9.0	18.0	11.7	6.6	16.0	18.0
Oldest and only	51.5	46.0	43.0	45.7	40.0	40.0	42.0

TABLE 13

	Census	Enuretics	Addicts	Child delinquent	Adult delinquent
Oldest	23.0	42.0	27.0	66.0	34.0
Middle	26.0	26.6	10.0	16.0	34.0
Youngest	23.0	21.0	24.0	12.0	32.0
Only	28.5	10.0	39.0	4.0	15.0
Oldest and only	51.5	52.0	64.0	70.0	49.0

and only groups in the population, we have 51.5%; in our child group, 46%; and in our adult group 43%. One notes the correspondence between the youngest census group and the children represented in families in the explosive and withdrawn groups, the manic and the schizoid groups. Here is a strong suggestion that the last position in the family is not especially significant and is in harmony with the proportions of the youngest children in the general population. This is suggested even in the consideration of our adult addict group and our juvenile enuretics (Table 13).

The variations from this norm among the juvenile delinquents and enuretics has already been explained. It would appear, therefore, quite obvious that being the youngest child in the family is not responsible for the explosive type of reaction, or the withdrawn personality. Nor is it to be deemed especially a factor in the development of manic or schizoid psychoses.

The manic and schizoid adult groups are quite in harmony with the census proportion of oldest children, even though we recognize that death changes the proportions of families during the course of years. The reason for a higher number of oldest children in the juvenile group has already been explained. If we add the oldest and only children, the census figures amount to 51.5%, in comparison with which we find the oldest and only children in the manic group to be 40%, in the schizoid group 42%, in the explosive group

45%, and in the withdrawn group 40%. It is evident not alone that the only children in both the adult and juvenile groups are below the number of children in the general population, but that the sum total of the oldest and the youngest children is below the totals of the crude distribution of these children in the general population.

One would expect the percentage of middle children to be above the average, and this fact is apparent in the explosive, the withdrawn, the manic, and the schizoid groups. The higher percentage of middle children among the adults and among the juvenile group probably results from the high mortality rates in infancy and childhood, and the mortality rate above age 40, which would tend to lower the number of surviving oldest children according to the laws of chance. It is also important to recognize that there is an increasing tendency to reduce the number of middle children in the population by purposeful reduction of the size of the family. According to the theory of family position as affecting style of living, the diminished birth rate should alter the character of the people. One questions such a conclusion.

In the Registration Area of the United States the percentage of the population under ten years of age was 33.5 in 1880, and only 23.7 in 1900. The number of children under five years to each thousand mothers between the ages of 15 and 59 years was 626 in 1850, and 474 in 1900. The distribution of age groups in the United States population has varied markedly from 1850. In 1850 the children under one year of age constituted 2.7% of the total population, while in 1920 it was 2.1%. In 1850 the children under age five constituted 15.1% of the population but only 10.9% in 1920. Children between the ages of 5 and 9 years constituted 14.0% of the population in 1850, and only 10.8% in 1920. Hence the total proportion of the population under 9 years of age in 1850 was 29.1%, as compared with 21.7%. This marked fall in population signifies a marked alteration in the position of children in families. This perhaps may be emphasized still more by a birth rate diminished to 21.4% in 1925 and an infant mortality rate of 71.7. The lower birth rate, the lower infant mortality rate, and contraception tend to produce a larger number of two-children families.

One may grant that there are individual environments for each child in the family, but this does not make it possible to conclude that these differences in environment are in themselves responsible

for marked alterations in personality. Goodenough and Leahy, from their experience, allege that oldest children tend to lack aggressiveness, lack in leadership and suggestibility; middle children are slightly more aggressive, very affectionate, usually gregarious, and usually not popular with other children; and there are no outstanding traits of youngest children. They report the only children to be more aggressive and more self-confident than any other grouping, as showing the greatest fondness for physical demonstration, and as likely to show more instability of mood.

It is frequently asserted that oldest children show the greatest proportion of undesirable extremes of behavior. This might in a sense appear reasonable because of the comparative inexperience of parents with their first-born children. It is also alleged that maladjustments arise for such children because they have difficulty in changing from an only child to a non-only child. This again is not borne out by our experience. We grant that at times the ordinal position of a child may be of considerable significance, but to generalize and to assert that the position in itself and the positional psychology results are responsible for profound changes does not appear to be supported by our experience. The similarities between our explosive and withdrawn children, as compared with the older groups of manic and schizoid adults, and their relation to the ordinal position distribution in the population tend to indicate a negative correlation.

The relation of the oldest child, in terms of success, have been studied by Galton, Cattell, Terman, Ogburn, and others, but even their results call for a careful evaluation of the numerous factors which enter into personality development. Granting that in our group of explosive and withdrawn children there is a larger number of older children, one is not able to state that their condition is due entirely to the fact of being first children. One cannot assume that it is due to the fact that they are the children upon whom the parents gain their first experience. Nor can it be stated dogmatically that the difficulties which brought them to the clinic would not have arisen had they been middle children. Possibly mothers are more disturbed by the activities of oldest children because of their strained nerves and the stress of the other one or more children. May not the small number of only children in our group be due likewise to greater maternal endurance when no example is set for other children to imitate? It is worthy of note that the percentages of oldest

children were equal in the explosive and the withdrawn types, despite the differences in personality.

Reynolds, discussing the emotional handicaps of some habit clinic children, regarded position of the child in the family to have only slight significance. Alfred Adler, who has laid the most stress upon the family constellation, states that the first-born tends to be conservative, the second-born more aggressive, and he mentions the fact that the "restless neurotics" are to a preponderating degree second-born children. The last-born he asserts to be "restlessly pushing forward, they surpass everyone by their initiative." Our experience fails to find any basis for this conclusion in the juvenile or adult groups studied. Our findings are thoroughly in harmony with the conclusions of L. L. and T. G. Thurstone, "We conclude, therefore, that the order of birth is not demonstrated to have any important relation to the development of a maladjusted or neurotic personality. This does not mean that being an only child, youngest child, oldest child, and so on, might not be an important contributory factor in the development of a neurotic personality. But our data do indicate that the question of birth order is not so universally important a consideration in mental hygiene as is sometimes believed."

Until there are more definite objective measures and tabulations of different races, of varying social and economic groups in the school population and in institutional population, and until there are more exact definitions of personal characteristics, it will be difficult to ascertain to what extent the ordinal position is an actual factor in establishing the style of living, in molding the essential traits of personality and in generating the mighty forces of character. Our studies suggest, however, that ordinal position is not an especially significant factor in fixing a personality type or in establishing definite forms of difficulty in adjustment or in determining dominant attitudes and responses in human relations.

*264 West Seventy-third Street
New York City*

ORDRE DE NAISSANCE ET COMPORTEMENT

(Résumé)

C'est une étude de 500 enfants et de 365 adultes en ce qui concerne l'ordre de naissance et les tendances générales du comportement. On a divisé les enfants en groupes représentant les types explosif et retiré de comportement et comparés aux adultes, maniaques et schizoïdes dans leur comportement psychiatrique. On a fait d'autres comparaisons des enfants avec énurésie

et les adultes avec penchants narcotiques et des comparaisons des jeunes et des adultes délinquents.

La distribution du pourcentage de la position ordinale des enfants et des adultes n'a montré nulle évidence de la domination d'une position spéciale parmi les groupes nommés explosif ou retiré, maniaque ou schizoïde. La distribution des enfants aînés, d'âge moyen, et plus jeunes a été parallèle à la distribution théorique pour la grandeur de famille et on l'a trouvée approximativement la distribution ordinale théorique des enfants dans la population générale.

Tandis que chaque enfant a un milieu individuel dans la famille, il est impossible de conclure que les différences de milieu par elles-mêmes sont responsables des changements marqués de personnalité. Il est difficile de déterminer l'importance de la position ordinale comme vrai facteur dans l'établissement de la manière de vivre, dans la formation des traits essentiels de la personnalité, et dans la génération de la grande force de caractère. L'étude suggère que, contrairement aux assertions d'Alfred Adler, la position ordinale ne possède nulle signifiante ni dans la fixation d'un type de personnalité ni dans l'établissement des formes définies de difficulté d'ajustement ni dans la détermination des attitudes et des réponses dominantes dans les relations humaines.

WILE ET NOETZEL

GEBURTSORDNUNG UND VERHALTEN

(Referat)

Es handelt sich um eine Untersuchung von 500 Kindern und 365 Erwachsenen mit Hinsicht auf die Geburtsordnung und den allgemeinen Charakter des Verhaltens. Die Kinder wurden in Gruppen explosiver und zurückgezogener Typen des Verhaltens eingeteilt und mit Erwachsenen, Manischen und Schizoiden in ihrem psychopathischen Verhalten verglichen. Es wurden ferner Vergleiche angestellt zwischen bettnässenden Kindern und dem Nikotin ergebenden Erwachsenen, ferner solche zwischen delinquenten Minderjährigen und Erwachsenen.

Die prozentuale Distribution der ordinalen Stellung der Kinder und der Erwachsenen wies keinerlei Beweise auf für das Vorherrschen einer der als explosiv oder zurückgezogen, manisch, oder schizoid benannten Gruppen. Die Verteilung der ältesten, mittleren und jüngsten Kinder parallelierte die theoretische Verteilung nach Grösse der Familie und ergab sich als ungefähr gleich der theoretischen, ordinalen Verteilung von Kindern der allgemeinen Bevölkerung.

Währenddem jedes Kind ein individuelles Milieu in der Familie hat, darf man daraus nicht schliessen, dass die Unterschiede des Milieus an sich für ausgesprochene Abweichungen der Persönlichkeit verantwortlich sind. Es ist schwer zu bestimmen, wie weit die ordinale Stellung ein wirklicher Faktor in der Festlegung eines Lebensstils, im Modulieren von Wesenszügen der Persönlichkeit und im Hervorbringen des Charakters ist. Die Untersuchung macht den Vorschlag, dass im Gegensatz zu Alfred Adlers Behauptungen, die ordinale Stellung weder Bedeutung hat in der Festlegung der Persönlichkeitstypen, noch in der Errichtung fester Formen von Anpassungsschwierigkeiten, noch in der Bestimmung vorherrschender Haltungen und Reaktionen in menschlichen Beziehungen.

WILE UND NOETZEL

THE MEASUREMENT OF ADULT INTELLIGENCE, AND THE REQUISITES OF A GENERAL INTELLIGENCE TEST*¹

From the Institute of Child Welfare, University of California

HERBERT S. CONRAD

I. THE MEASUREMENT OF ADULT INTELLIGENCE

Any one who goes into the field to secure intelligence-test records on unselected adults invariably runs into a simple, exasperating difficulty; viz., a considerable percentage of the subjects decline to be tested. It is extremely difficult, without offering cash awards, or disguising one's purpose, or spending much time in gaining rapport, to persuade an ordinary unselected adult to take a mental test. Thus Willoughby (69) reports that refusals of adults to cooperate in his mental-testing program amounted to 50%. Freeman et al (20, pp. 141-142) obtained only 180 and 256 formal intelligence records from 394 and 401 foster-mothers and foster-fathers, respectively. While non-cooperation may not have been the only factor operating in Freeman's case, it nevertheless appears to be an important one. In contrast, Jones (30) succeeded in obtaining a very high percentage of cooperation from rural adults; but this involved much greater expense and special preparation than would ordinarily be feasible. Burks (7, pp. 240-241), too, has been exceptionally successful; we are not familiar with the details of her administrative technique. Of course, the Army psychologists, under very special conditions, tested large numbers of adults (70); and Murchison (37) has successfully tested prisoners. Neither soldiers nor prisoners, however, can be relied upon to furnish a fair sample of adult intellect at any age level.

As a matter of fact, securing adult cooperation in intelligence tests is so ticklish a matter that until recently it has seldom even

*Recommended by Harold E. Jones, accepted for publication by Carl Murchison of the Editorial Board, and received in the Editorial Office March 3, 1930.

¹The writer is indebted to Professor Harold E. Jones of the University of California for suggesting this paper and for a number of contributions to its contents.

been attempted on any extensive scale (54). Thus, a careful summary by Burks (8) of the psychological literature on heredity reveals an abundance of studies of intelligence-resemblances between siblings; but a complete absence (prior to 1928²) of studies of objectively measured intelligence resemblances between *parent* and child! Nor has this state of affairs been due to any lack of interest in the problem of parent-child relationship. On the contrary, the fairly numerous studies of the relation between intelligence of children and such factors as education of parents, economic status of parents, and *ratings* of parents' intelligence (8)—all alike point to a very strong interest.

Evidence of the adult resistance to present intelligence tests is easy to find in various other fields. For example, home scales and cultural scales have frequently been used, in which parental intelligence is nowhere directly included (7, pp. 231-235; 25). Again, in the very careful and detailed work of many child welfare institutions, an increasingly full continuous record is being kept of children's developmental histories, home conditions, parental attitudes, etc.; but practically never is any provision made for the record of objective intelligence tests of the adult members of the children's households. The importance of such information in the study of, say, adolescent and post-adolescent maladjustments is clear. Even many employers pause to deliberate before asking their prospective employees to take an undisguised, formal intelligence test; nor is the outcome of such deliberation always in favor of administering the test (68). Here, as in the previous instances, the prime causes for the absence of adult testing are not considerations of utility, but of difficulty, of cost, and of possible (not to say probable) *malapprochement*.

The unwillingness of adults to be tested is not limited to the less intelligent classes, nor to the usual type of intelligence test. M. B. Blanchard, a psychiatric social worker, reports the same reluctance, in the case of middle-class adults, to take such a test as the Pintner Non-Language (41).³ Whether this reluctance would occur in equal measure for performance tests we are not able to say. In any case, performance tests do not, to date, correlate sufficiently well with mental tests to serve as an adequate substitute (2). Performance tests applied to normal individuals may supplement, but can scarcely replace mental tests.

²The studies of parent-child relationship published in 1928 include those of Willoughby, Freeman, Jones, and Burks, already mentioned.

³Unpublished data reported to the writer.

To inquire here into the causes of the adult aversion to mental tests would carry us too far afield into hypothesis and speculation. The practical facts are, first, that this adult aversion does not appear to be noticeably abating; and, secondly, that it seems impossible to construct a *formal* intelligence test which will, without questionable "sugar-coating" or highly expensive administrative technique, escape the displeasure of the ordinary, unselected adult.

But it is precisely this ordinary, unselected adult whom the psychologist most frequently wants to discuss. Not having precisely the data needed, the customary procedure of the psychologist so far has been to draw conclusions about the average *homo sapiens* from experimental evidence which has been secured from groups of school children, or college students, or army draftees, or psychopathic patients or institutional inhabitants with one sort of sampling peculiarity or another. Indeed, the animal psychologist, accused by the human psychologist of extending his conclusions much beyond his experimental data might well ask whose privilege it is to throw the first stone.

The practical problem remaining for intelligence testers, then, is either to alter the attitude of adults to current intelligence tests; or to alter the tests to suit the attitude of the adults. The latter alternative—viz., the construction of a good general intelligence test which will be *acceptable to the persons to be tested*—appeals to us as being more practical.

The virtue of acceptability, however, may not be completely attainable without the sacrifice of other desiderata. Because of the recent activity in the construction of acceptable tests for adults (12, 11, 54), we venture to present, briefly, a list of the requirements of a good general intelligence test for adults. With these requirements explicitly in mind, new tests may be constructed with more balanced emphasis upon the various desirable features; and both the new tests and the old may perhaps be evaluated with greater insight, and less narcotic satisfaction.

The various desiderata follow. No attempt can be made, in this short article, elaborately to justify and evaluate each individual item (38), nor to trace interesting analogies in the testing of children, infants, and animals.

II. DESIDERATA OF A VALID GENERAL INTELLIGENCE TEST

1. The test should have a *definite zero-point* (60, 61).

2. Scores should be obtained in terms of a scale made up of equal units:⁴ or, at any rate, the scores obtained should be susceptible to techniques available for scaling of a test into equal units (60, 61a).

3. The test should measure *innate*⁵ *intellectual capacity*, rather than acquired intellectual ability (19); or, if the measurement is of both, it should, ideally, be possible quantitatively to separate the measure of capacity from the measure of ability. At present, awaiting an authenticated physiological (63) or anatomical (22, 53) method of measuring intelligence which yields sufficiently high validity coefficients (and which, in comparison with mental tests, is relatively insusceptible to a variety of uncontrollable embryological, physiological, environmental, and emotional factors),—at present, we are forced to *infer* innate capacity from measured acquired ability. Intellectual ability, of course, is influenced by such spurious factors as unequal opportunity, and unequal application (initiative, enterprise, and persistence). An intelligence test totally unaffected by these factors and the others mentioned later in this article, would be difficult to conceive. Probably the best we can hope for, then, is a very reliable ability score to which corrections for disturbing, non-standard factors may be made.

3a. The sample of tasks set in the test must be adequate to measure *general* intellectual capacity [unless a single type-task is found which correlates sufficiently well with a suitable composite of tasks (33, 56, 60)].

3b. Language and vocabulary factors should not be overweighted; non-verbal, or less-verbal intelligence should be duly taken care of (44, 52). Similarly, rote memory should not be measured at the expense of critical or suggestive or organizing talent (9, 48, 65).

⁴The Stanford-Binet Test, e.g., according to which the average adult intelligence is inferred to be about the same as the average 13-16-year-old child (17, 18, 39), and according to which the IQ drops with age (4; 20, p. 109; 26), appears to suffer from expanded units in its upper range (59). For tests other than the Stanford-Binet, see reference (60), Chapters VII, "The Transformation of the Scores of Standard Examinations into Terms of Scales with Equal Units."

⁵Such as may object to the use of the word "innate" in connection with intelligence, believing that intelligence is largely an environmental and not a hereditary trait, are referred to the recent researches on this subject in reference (64).

3c. The test should measure the same thing⁶ (viz., innate intellectual capacity) *without change over a long span of years* (8a), in order to permit year-to-year, and generation-to-generation comparisons. (The latter is especially necessary in studies of heredity.)

3d. The variations upward or downward of an individual's test scores over a period of years should represent true variations, and not test artifacts due to environmental fluctuations (7, p. 306; 20, p. 49), or to test unreliability, or to change in test material with age.

3e. In the present state of controversy, the test(s) should furnish the data from which a measure of "*g*," "*s*," and "group factors" may be derived (29, 33, 55), as well as data for measures of level, range, and speed (60).

4. The test and individual test-items should possess a high, uniform, known degree of statistical reliability [factors affecting the reliability of psychological tests applied to humans have been listed by Symonds (58)].

5. The criteria of intelligence (or, in other words, the practical definition of intelligence) against which the test and the individual test items have been (or are to be) validated, should be explicitly stated (33, 55, 60, 65). These criteria should be adequate, and themselves reliable. Non-verbal and long-time criteria should not be unduly sacrificed for the convenience of verbal or short-time ones (70, p. 452).

6. The test and the individual test-items should possess a high, (preferably) uniform, and known degree of validity with respect to stated criteria. If the elements of the test are not equally reliable or valid, they should be appropriately weighted (32; 70, p. 452; 67).

7. The items in the test, aside from being uniformly reliable and valid, should have a low intercorrelation (21). Similarly, the individual tests of a test battery should also have a low intercorrelation.

8. *Factors mainly biologic, social, and ethnic:*

The test, both in respect to directions, content, and mode of subject's response, should be equally applicable, reliable, and valid

⁶Test 8 of the Army Alpha (information test) (70, pp. 227, 234) is an excellent illustration of a test which requires revision every decade or so. It is, of course, difficult to be sure (especially in an *individual* case) that the revision of any test is the same as the original test in respect of difficulty and the ability which is measured.

for individuals differing in sex (31, 35), age⁷ (4; 10; 20, p. 109; 26, 28, 59), sensory equipment [e.g., deaf vs. hearing (43)], schooling (1, 14), socio-economic status (7, 20), and culture complex [in the anthropologist's sense of the word (16, 34, 57)].

When such broad applicability is not possible, the use of the test becomes, for most theoretical problems, either sharply limited or subject to "corrective factors." These limitations and corrective factors should in every case be determined and stated both qualitatively and quantitatively. It cannot be over-emphasized in this connection, that a "correction" is practically always an *average* figure, which usually does not apply with anything like strict validity to an *individual* subject. For this reason, it is probably best to construct tests with a view to maintaining all necessary corrections at as low a level as possible;⁸ in this way the very possibility of large error in the application of the correction to individual subjects is eliminated.

The remarks in the preceding paragraph apply with equal force to the factors mentioned below under 9.

9. *Factors mainly personal:*⁹

a. The test, both in respect to directions, content, and mode of subject's response, should be equally applicable, reliable, and valid for individuals differing widely in intelligence. The difficulties involved in devising a test at once applicable to a young idiot and an adult genius are obvious. The ideal complete test would be composed of a great many items of known scale value, from which an appropriate selection could be made to suit populations of various degrees of capacity and homogeneity (61*b*).

b. Differences of ability or facility in reading and writing, and differences in sensori-motor speed or acuity should have a minimal (or a known, or measurable) effect upon test results (50) (*except insofar as these differences are themselves due to differences in intelligence; this qualification obviously applies also to factors mentioned below under c, d, e, f, and 12.*)

c. Differences in physical condition (27), temperament (46),

⁷All tests requiring reading, for example, are inapplicable to most children below ten years.

⁸This caution would not be necessary, if the disturbing factor were accurately measured for each individual subject, and if the influence of the disturbing factor were highly uniform and exactly known.

⁹For convincing evidence as to the reality and importance of these factors, attention is directed especially to Chapter 3 of (66).

and mood (5, 29) should have only a minimal (or a known, or measurable) effect upon the test results. Both directions and test content should be specifically designed to discourage undue hesitancy, timidity, nervousness, or excess confidence.

d. Differences in the standards of speed and accuracy which the test induces the examinee to set for himself should be either mutually compensatory, or be at a minimum and have only a minimal (or a known, or measurable) influence upon the test result.

e. Differences in conditions of temperature, humidity, ventilation, illumination, time of day and writing instrument used (where response is written) should have only a minimal (or a known, or measurable) effect upon the test results.

f. Environmental and personal accidents [such as noise (62), conversation, sneezing, breaking of a pencil, defective test booklet, imperfections of timing, etc.] should be sufficiently controllable through the conditions of administration, so that they will have only a slight effect upon the test results.

With respect to the items listed under 9b-e, the ideal result would be such that a test would be *untaken, or obviously incomplete and invalid*, as soon as the damaging influence of the various factors mentioned becomes of any importance. But, on the other hand, an unusual degree of *freedom* from such damaging influences should not be capable of raising the subject's test score appreciably—unless the degree of spurious rise is either measurable or known.

10. The prescribed satisfactory conditions of administration should be practical, convenient, flexible, and easily attained at low cost. The test should be susceptible to both group and individual administration, as well as to self-administration by mail (11).

11. The directions for the test should be simple, clear, self-explanatory, and effectively illustrated.

12. The directions and test should be intrinsically interesting, in order to elicit and maintain a uniform maximum of serious effort and cooperation, and in order to prevent the intrusion of irrelevant competing ideas.

13. The test should be *acceptable to the persons to be tested*. No test which fails to carry with it a suitable degree of interest, motivation, and acceptability can be expected to serve as a useful index of the subject's intelligence. This requirement of acceptability probably means, for adults, that the intelligence test will have to be somewhat disguised. Directions and test content must be specially selected

and modified to call forth a continuously favorable response from the subject; words or test items which provoke an unpleasant emotional tone should be avoided; printed tests should be attractive and interesting in appearance, preferably illustrated and in colors; oral tests should, if necessary, provide some interesting introduction and running comment. Especially in group tests, the whole administrative procedure must be very carefully arranged; for acceptability in adult testing can be won or lost by the manner in which a test is presented (13).

14. Other things being equal, the less time required of tester and testee, the better. Especially in group tests, administrative factors may require severe limitation of the time allowance (13), even if such limitation is unnecessary from the point of view of test validity (60).

15. The test content should preferably be such that there is plenty of material readily available for retesting with alternate forms of equal difficulty, reliability, validity, and applicability.

16. Practice effects should have only a minimal (or measurable) influence upon the test result (16, pp. 81-83; 24). If, however, there is danger of coaching (see below), practice effects should be *minimal*. (Of course, one way to minimize practice effects is to give to all subjects a suitable preliminary practice form.)

17. Coaching (23) should be a negligible factor, both through the availability of many alternate forms, and the impotence of practice to work differentially under the conditions of administration (see 16 above).

18. The test should be such that under test conditions there will be stimulation to individual, non-cheating effort, together with a lack of opportunity for cheating (13). Cheating, if sufficiently extensive to affect test results significantly, should become detectable (3).

19. Chance factors should be at a minimum not only for the *group*, but for the *individual* (this is a criticism of many true-false and multiple-choice tests). Ambiguities, catch questions and over-fine subtleties should be avoided. The printed tests (or the oral presentations) should be uniform, and such that the reply is uninfluenced by the mechanics of presentation (36).

20. The scoring of the test should accord proper, weighted values to every test item or test element which contributes toward prediction of the criterion (40). The method of scoring should

take into account both the nature of the directions (e.g., "guess" vs. "do not guess"), and the nature of the test items (47, 51). Scoring should be objective, quick, accurate, readily susceptible to checking, and inexpensive (42).

21. The expense of administration, scoring, and test material or apparatus should be at a minimum. Test material should be readily and cheaply replaceable in case of damage or loss.

22. The record of the subject's responses should have a high degree of physical permanence. Ideally, each subject should record his responses in such manner that scoring, determination of test reliability, and other statistical computation could readily be done by suitable machines.

23. A set of *reliable* norms of the test, both for the total population and for different subclasses of the total population is essential. If certain biologic, social, ethnic, or personal factors influence the test results, these limitations of the test should be emphasized. Ideally, norms which make allowance for disturbing factors (see especially 8 and 9 above), or for combinations of such disturbing factors, should be provided.

24. The test should, ideally, be more than simply an instrument to measure an individual's total general intelligence. Without loss of efficiency in its main purpose, the test should attempt also to locate the individual's specific fields of intellectual strength and weakness, and should give at least some tentative indication with reference to procedures designed to capitalize the particular strength, or to mitigate the particular weakness. It has also been urged that the intelligence test should, in addition, be at least suggestive of *unusual* strength or deficiencies in temperamental or character traits, wherever such exist (6, 45).

III. SUMMARY AND CONCLUSIONS

Present intelligence tests are inadequate for testing an unselected sample of adults, since they cannot be used except by questionable sugar-coating, or special persuasion, or the offering of extraneous rewards. This accounts for the very few studies reporting objective measures of adult intelligence. The practical problem now remaining for intelligence testers is to construct a good general intelligence test *which will be acceptable to the persons to be tested*.

A list of desiderata of a valid adult intelligence test has been presented, with respect to which both new and old intelligence tests

may be evaluated. The desiderata, however, are so numerous and so difficult to attain, that probably not a single present intelligence test (or even, perhaps, battery of tests) can escape the verdict of being pitifully deficient. The list of desiderata should, therefore, be regarded as an ultimate ideal, although this ideal is perhaps too high to form the immediate goal of present-day test construction.

To the practical reader, applying a purely pragmatic, momentary standard, it may seem that in several respects the list of desiderata merely strains at many gnats. To such a reader it is suggested that he thoroughly study and digest the formula for the standard error of estimate (21, pp. 183, 237). It must be remembered that a reduction of the standard error of estimate by 70% requires a correlation of no less than .95. This is higher than the reliability (or *self-correlation*) of many intelligence tests in homogeneous groups.

Nothing is more certain than that no science can rise above the merits of its measurements. Validity coefficients which, in groups homogeneous both as to intellectual maturity and practical social or economic characteristics, fall below .90—.95 (failing to reduce the standard error of estimate by as much as 65-70%) can hardly have more than an emotionally stimulating or a dubiously suggestive value in any rigorous scientific theory. In practice, however, a coarse, qualitative measure may often serve virtually as well as a fine.

It will occur to the practical reader that much of the criticism implicit in this article has been destructive rather than constructive. However, a series of other articles present in detail some intelligence tests which, both in test content and administrative technique, have at least achieved the fundamental virtue of acceptability (11, 12, 13).

REFERENCES

1. BAGLEY, W. C. *Determinism in education*. Baltimore, Md.: Warwick & York, 1925. Pp. 194.
2. BAYLEY, N. Performance tests for three-, four-, and five-year-old children. *Ped. Sem.*, 1926, **33**, 435-454.
3. BIRD, C. The detection of cheating in objective examinations. *School & Soc.*, 1927, **25**, 261-262.
4. BLANCHARD, M. B. The applicability of the Stanford revision of the Binet-Simon scale of intelligence to a typically rural group of children. M. A. thesis, Columbia Univ., 1927.
5. BRIDGES, K. M. B. Critical notes on mental tests for children of pre-school age. *Ped. Sem.*, 1927, **34**, 38-44.
6. BROWN, W. M. A study of the caution factor and its importance in intelligence test performance. *Amer. J. Psychol.*, 1924, **35**, 368-386.

7. BURKS, B. S. The relative influence of nature and nurture upon development. *27th Yrbk. Nat. Soc. Stud. Educ.*, Pt. 1, 1928, 219-316.
8. ———. A summary of literature on the determiners of the intelligence quotient and the educational quotient. *27th Yrbk. Nat. Soc. Stud. Educ.*, Pt. 2, 1928, 248-353.
- 8a. ———. When does a test measure the same functions at all levels? *J. Educ. Psychol.*, 1930, **21**, 616-620.
9. CANNON, J. G. An enquiry into the relative values of the inventive and selective forms of group tests of mental capacity. *Australasian J. Psychol. & Phil.*, 1926, **4**, 141-149.
10. CONRAD, H. S. General information, intelligence, and the decline of intelligence. *J. Appl. Psychol.*, 1930, **14**, 592-599.
11. CONRAD, H. S., & HARRIS, D. The free association method and the measurement of intelligence. *Univ. Calif. Publ. Psychol.*, 1931, **5**, 1-45.
12. CONRAD, H. S., & JONES, H. E. Psychological studies of motion pictures. III. Fidelity of report as a measure of intelligence. *Univ. Calif. Publ. Psychol.*, 1929, **3**, 245-276.
13. ———. Psychological studies of motion pictures. IV. The technique of mental test surveys among adults. *Univ. Calif. Publ. Psychol.*, 1929, **3**, 277-284.
14. DAVIS, R. A. Some relations between amount of school training and intelligence among negroes. *J. Educ. Psychol.*, 1928, **19**, 127-130.
15. DEARBORN, W. F., ANDERSON, J. E., & CHRISTIANSEN, A. O. Construction tests of mental ability. *J. Educ. Psychol.*, 1916, **7**, 445-458.
16. DODD, S. C. International group mental tests. Princeton, N. J.: Princeton Univ., 1926.
17. DOLL, E. A. The average mental age of adults. *J. Appl. Psychol.*, 1919, **3**, 317-328.
18. ———. The growth of intelligence. *Psychol. Monog.*, 1921, **29**, No. 131. Pp. vi+130.
19. FREEMAN, F. N. Mental tests. Boston: Houghton Mifflin, 1926. Pp. ix+503.
20. FREEMAN, F. N., HOLZINGER, K. J., & MITCHELL, B. C. The influence of environment on the intelligence, school achievement, and conduct of foster children. *27th Yrbk. Nat. Soc. Stud. Educ.*, Pt. 1, 1928, 102-217.
21. GARRETT, H. E. Statistics in psychology and education. New York: Longmans Green, 1926. Pp. 330.
22. GARRETT, H. E., & KELLOGG, W. N. The relation of physical constitution to general intelligence, social intelligence, and emotional instability. *J. Exper. Psychol.*, 1928, **11**, 113-129.
23. GILMORE, M. E. Coaching for intelligence tests. *J. Educ. Psychol.*, 1927, **18**, 119-121.
24. GUNDLACH, R. The effects of practice on the correlations of three mental tests. *J. Educ. Psychol.*, 1926, **17**, 387-401.
25. HEILMAN, J. D. The relative influence upon educational achievement of some hereditary and environmental factors. *27th Yrbk. Nat. Soc. Stud. Educ.*, Pt. 2, 1928, 35-65.

26. HIRSCH, N. D. M. An experimental study of the East Kentucky Mountaineers. *Genet. Psychol. Monog.*, 1928, **3**, 183-244
27. HOEFER, C., & HARDY, M. C. The influence of improvement in physical condition on intelligence and educational achievement. *27th Yrbk. Nat. Soc. Stud. Educ.*, Pt. 1, 1928, 371-388.
28. HSIAO, H. H. The performance on the Army Alpha as a function of age. M. A. thesis, Columbia Univ., 1927.
29. HULL, C. L. Aptitude testing. Yonkers, N. Y.: World Book Co., 1928. Pp. xiv+535.
30. JONES, H. E. A first study of parent-child resemblance in intelligence. *27th Yrbk. Nat. Soc. Stud. Educ.*, Pt. 1, 1928, 61-72.
31. JONES, H. E., CONRAD, H. S., & HSIAO, H. H. The growth and decline of intelligence. (Not yet published.)
32. KELLEY, T. L. Statistical method. New York: Macmillan, 1923. Pp. xi+390.
33. ———. Crossroads in the mind of man: a study of differentiable mental abilities. Stanford University, Calif.: Stanford Univ. Press, 1928. Pp. 245.
34. KLINEBERG, O. An experimental study of speed and other factors in "racial" differences. *Arch. Psychol.*, 1928, **15**, No. 93. Pp. 111.
35. LINCOLN, A. Sex differences in the growth of American school children. Baltimore, Md.: Warwick & York, 1927. Pp. xii+189.
36. MATHEWS, C. D. The effect of position of printed response words upon children's answers to questions in two-response types of tests. *J. Educ. Psychol.*, 1927, **18**, 445-457.
37. MURCHISON, C. Criminal intelligence. Worcester, Mass.: Clark Univ., 1926. Pp. 291.
38. OTIS, A. S. Scale for rating tests. Test Service Bulletin No. 13. Yonkers, N. Y.: World Book Co., 1926.
39. ———. Some logical aspects of the Binet scale. *Psychol. Rev.*, 1913, **23**, 129-152; 165-179.
40. PETERSON, J. Intelligence and its measurement. *J. Educ. Psychol.*, 1921, **12**, 198-201.
41. PINTNER, R. A non-language group intelligence test. *J. Appl. Psychol.*, 1919, **3**, 199-214.
42. ———. A new intelligence test for survey purposes. *Teach. Coll. Rec.*, 1927, **29**, 18-23.
43. ———. A mental survey of the deaf. *J. Educ. Psychol.*, 1928, **19**, 145-151.
44. PINTNER, R., & PATERSON, D. G. A scale of performance tests. New York: Appleton, 1917. Pp. 13.
45. POFFENBERGER, A. T., & CARPENTER, F. L. Character traits in school success. *J. Exper. Psychol.*, 1924, **7**, 67-74.
46. REGENSBURG, J. Social implications of mental testing. *Family*, 1927, **7**, 295-300.
47. RICHARDS, O. W. High test scores attained by subaverage minds. *J. Exper. Psychol.*, 1924, **7**, 148-156.
48. ROBACK, A. A. Subjective tests vs. objective tests. *J. Educ. Psychol.*, 1921, **12**, 439-444.

49. ROGERS, A. L., DURLING, D., & MCBRIDE, K. The effect on the intelligence quotient of change from a poor to a good environment. *27th Yrbk. Nat. Soc. Stud. Educ.*, Pt. 1, 1928, 323-331.
50. ROOT, W. T. Two cases showing marked change in I.Q. *J. Appl. Psychol.*, 1921 **5**, 156-158.
51. RUCH, G. M., & DEGRAFF, M. H. Corrections for chance and "guess" vs. "do not guess" instructions in multiple response tests. *J. Educ. Psychol.*, 1926, **17**, 368-375.
52. SAER, P. J. An inquiry into the effect of bi-lingualism upon the intelligence of young children. *J. Exper. Ped.*, 1922, **6**, 232-240; 266-274.
53. SHELDON, W. H. Morphological types and mental ability. *J. Person. Res.*, 1927, **5**, 447-466.
54. SNEDDEN, D. S. A study of disguised intelligence tests. *Teach. Coll. Contrib. Educ.*, 1928, No. 291. Pp. 48.
55. SPEARMAN, C. E. The abilities of man: their nature and measurement. New York: Macmillan, 1927. Pp. xxiii+415.
56. ———. The uniqueness of "g." *J. Educ. Psychol.*, 1929, **20**, 212-216.
57. SQUIRES, P. C. A universal scale of individual performance tests. Princeton, N. J.: Princeton Univ. Press, 1926. Pp. 158.
58. SYMONDS, P. M. Factors influencing test reliability. *J. Educ. Psychol.*, 1928, **19**, 73-87.
59. TERMAN, L. M., et al. Genetic studies of genius. Vol. I. Mental and physical traits of a thousand gifted children. Stanford University, Calif.: Stanford Univ. Press, 1925. Pp. xvi+648.
60. THORNDIKE, E. L. The measurement of intelligence. New York: Teach. Coll., Bureau of Publs., 1926. Pp. xxvi+616.
61. THURSTONE, L. L. The absolute zero in intelligence measurement. *Psychol. Rev.*, 1928, **35**, 175-197.
- 61a. ———. A method of scaling psychological and educational tests. *J. Educ. Psychol.*, 1925, **16**, 433-451.
- 61b. ———. The scoring of individual performance. *J. Educ. Psychol.*, 1926, **17**, 446-457.
62. TINKER, M. A. Intelligence in an intelligence test with an auditory distractor. *Amer. J. Psychol.*, 1925, **26**, 467-468.
63. TRAVIS, L. E. The correlation between intelligence and speed in conduction of the nerve impulses in a reflex arc. *Science*, 1928, **67**, 41-43.
64. [VARIOUS.] Nature and nurture. *27th Yrbk. Nat. Soc. Stud. Educ.*, 1928.
65. ———. Intelligence and its measurement: a symposium. *J. Educ. Psychol.*, 1921, **12**, 123-147; 195-216.
66. WENTWORTH, M. M. Individual differences in the intelligence of school children. *Harvard Stud. Educ.*, 1926, No. 7. Pp. 158.
67. WEST, P. V. The significance of weighted scores. *J. Educ. Psychol.*, 1924, **15**, 302-312.
68. WIEMAND, C. One industry's attitude toward selection by mental levels. *J. Educ. Psychol.*, 1924, **15**, 180-182.

69. WILLOUGHBY, R. R. Family similarities in mental-test abilities. *Genet. Psychol. Monog.*, 1927, 2, 236-277.
70. YERKES, R. M. Psychological examining in the United States Army. *Mem. Nat. Acad. Sci.*, 1921, 15. Pp. vi+890.

*University of California
Berkeley, California*

LA MESURE DE L'INTELLIGENCE DES ADULTES, ET LES QUALITÉS REQUISES D'UN TEST GÉNÉRAL DE L'INTELLIGENCE

(Résumé)

Il est très difficile, sans offrir de l'argent, ou cacher son dessein, ou passer beaucoup de temps à gagner la coopération, de persuader à un adulte ordinaire, non choisi, de subir un test mental étalonné. Cette difficulté de tester l'intelligence des adultes est un handicap sérieux dans les études de l'hérédité, des relations entre les parents et les enfants, de la croissance et de la décroissance de l'intelligence, et de l'intelligence des divers groupes sociaux. Pour cette raison, on a formulé récemment nombre de tests d'intelligence, dont la valeur propre est leur *acceptabilité* aux adultes. Malheureusement, ces tests ne possèdent pas l'acceptabilité sans le sacrifice (ou du moins le non-acquisition) d'autres qualités désirables. Pour qu'on puisse construire des tests nouveaux pour les adultes avec une compréhension plus balancée, et qu'on puisse évaluer les tests anciens et nouveaux avec plus de justesse et moins de satisfaction narcotique, on présente une compilation des plus importantes qualités désirées d'un test d'intelligence. Les qualités désirées sont cependant si nombreuses et si difficiles à atteindre, qu'il n'y aura un seul test d'intelligence employé actuellement qui puisse s'appeler sans défauts. On ajoute une bibliographie de 70 titres, comme orientation pour un lecteur qui désire considérer quelques-unes des questions en plus grand détail.

CONRAD

DIE MESSUNG DER INTELLIGENZ ERWACHSENER, UND DIE ANFORDERUNGEN EINES ALLGEMEINEN INTELLIGENZTESTES

(Referat)

Es ist ausserordentlich schwer, ohne Barentschädigungen, oder Verhüllung der Absicht, oder ohne Verlust von viel Zeit zur Herstellung eines Reports, einen gewöhnlichen, nicht erlesenen Erwachsenen zu bestimmen, sich einem psychotechnischen Normaltest zu unterziehen. Die Schwierigkeit, die Intelligenz Erwachsener zu messen, ist ein schweres Hindernis im Studium der Vererbung, der Eltern-Kinder-Beziehungen, der Intelligenzzu- und -Abnahme, und der Intelligenz verschiedener gesellschaftlicher Gruppen. Darum wurde kürzlich eine Anzahl von Intelligenztesten erfunden, deren besonderes Vorzugsmerkmal es ist, Erwachsenen annehmbar zu sein. Leider erlangen diese Tests die Eigenschaft, annehmbar zu sein, nicht ohne andere wünschenswerte Qualitäten zu opfern (oder wenigstens nicht zu er-

reichen). Damit neue Tests für Erwachsene mit besserem Verständnis aufgestellt, und alte und neue Tests mit grösserer Gerechtigkeit und weniger narkotischer Befriedigung bewertet werden können, wurden die wichtigen Anforderungen an einen Intelligenztest dargestellt. Die Erfordernisse sind aber so schwer zu erreichen, dass wahrscheinlich nicht ein einziger der bestehenden Intelligenzteste dem Urteil entgeht, erbärmlich unzureichend zu sein. Der Arbeit folgt ein Literaturverzeichnis von 70 Titeln, um Lesern, die irgend eine Frage genauer betrachten wollen, eine Orientierung zu ermöglichen.

CONRAD

ORGANIC INFERIORITY AND THE INFERIORITY ATTITUDE*¹

From the Department of Psychology, University of Minnesota

HANNA F. FATERSON

"All manifestations of neuroses and psycho-neuroses are to be traced back to organ inferiority, to the degree and nature of central compensations that have not yet become successful, and to the appearance of compensation disturbances" (3, p. 316). In this statement, Adler summarizes his views on the relation of the inferiority attitude to organ inferiority. Both the organic and the "psychic" inferiority, as described by Adler in the above and in other publications (1, 2), are, of course, derived largely from pathological cases that have come to him for psychiatric treatment. The question arises whether this coexistence of organic inferiority and feelings of inadequacy will be found in the "normal" population—a population which, at least, has not been selected on the basis of seeking psychiatric advice. An inquiry into this problem was made possible by the use of the Heidebreder Rating Scale of Personal Traits, which was standardized on college students, and which includes the traits considered by Adler to be characteristic of the inferiority attitude (5, 4). This inquiry was undertaken with two problems in mind. In the first place, there is the theoretical question whether Adler's hypothesis as to the organic basis of feelings of inferiority holds for the normal population. In the second place, an inquiry into this relationship would offer an opportunity to validate, from one point of view, the measuring device which has its roots in the concept of

*Recommended by John E. Anderson, accepted for publication by Carl Murchison of the Editorial Board, and received in the Editorial Office, March 28, 1930.

¹This article contains a part of the material included in a Doctor's Thesis. The complete investigation of which the present article is a part is on file at the University of Minnesota library. The writer wishes to take this opportunity to express her appreciation to Dr. Edna Heidebreder who directed the investigation throughout its course, and who gave invaluable suggestions in the organization and interpretation of the material. Thanks are also due to Professor D. G. Paterson for permission to use the material collected by the University on the entering freshmen, and for helpful suggestions and criticisms.

organic inferiority and should, on theoretical grounds, bear some relation to it.

The general plan of the present study was to analyze the data of the physical examinations of university students whose scores on the Inferiority Attitude Rating Scale² were available, and to determine whether any relationship exists between the two aspects of Adler's clinical picture.

The physical data used were obtained from the Students' Health Service of the University of Minnesota.³ These examinations are a part of the routine of admission to the University, no student being exempt; the sampling used is, therefore, representative of the university population at entrance. Furthermore, since the data were collected by the physicians prior to and independently of their use for a study of temperament, there is little danger of a bias in that direction entering into the medical diagnosis. The use of a standard card for recording the results of the examinations insures a certain uniformity in recording the symptoms found.

In addition to the medical records, in the case of women, there was also available a set of health questionnaires, which the students themselves were asked to fill out. For the present purpose this, of course, is not treated as a valid index of the students' condition, but simply as an indication of their subjective estimate of their health—a very different proposition. These data will be referred to as self-estimates of health and will be treated separately.

The health items finally selected for the study are summarized in Tables 1-3. A more adequate explanation of the meaning of the items used is not possible here for lack of space.⁴

It may seem from looking over the list of physical imperfections as they appear in Tables 1-3 that they are for the most part but superficial minor complaints whose sum total never could make up the fundamentally inferior constitution. However, in reading Adler, one gets the impression that the inferior quality or condition of any given organ is important not so much in itself, but as an indication

²The Heidebreder scale will be referred to as the Inferiority Attitude (or I.A.) Rating Scale, referring to the origin of the material.

³For permission to use these records the writer is indebted to Dr. Diehl, Director of the Students' Health Service, and to Miss Kisson, Chairman of the Department of Physical Education for Women, University of Minnesota.

⁴More detailed explanations, together with the consideration of the debatable questions as to what, in any given case, constitutes a defect, will be found in the original thesis, on file in the library of the University of Minnesota.

TABLE 1

Table showing the average I.A. scores of groups of women for whom a given physical characteristic or defect is recorded, and the comparisons of these scores with control groups.

Classification of difficulties recorded†	No.	Inferiory attitude score for group			Comparison with the rest of the women exclusive of a given group			Comparison with the group for whom no positive findings are recorded			Average of 5c and 6c
		Av.	S.D.	Diff.	P.E.	Diff.	P.E.	Diff.	P.E.	Diff.	
Column number 1	2	3	4	5a	5b	5c	6a	6b	6c	7	
1. Menstr. disord. (B)	119	45.6	39.6	11.9	2.9	4.0	21.3	3.7	5.7	4.8	
2. Fallen arches	79	47.4	37.2	11.7	3.1	3.8	23.1	4.2	5.7	4.7	
3. Hands (tremor, etc.)	161	42.5	37.5	9.9	2.5	4.0	18.2	3.5	5.2	4.6	
4. Orthopedic gym.	109	45.9	44.5	10.7	3.1	3.4	21.6	4.1	5.3	4.3	
5. Strength of muscles	110	44.1	37.8	8.6	2.8	3.1	19.8	3.8	5.5	4.3	
6. Heart murmurs	82	46.8	43.6	11.1	3.5	3.2	22.5	4.3	5.2	4.2	
7. Thyroid enlarged	181	42.1	37.9	6.8	2.4	2.8	17.8	3.1	5.7	4.2	
8. Posture, poor	207	41.9	39.9	7.4	2.4	3.0	17.6	4.3	4.1	3.5	
9. Rea. to exercise	71	44.6	44.9	8.3	3.8	2.2	20.3	4.6	4.4	3.3	
10. Nasal obstruction	145	40.8	40.7	5.7	2.6	2.2	16.5	4.4	3.7	2.9	
11. Tonsils defective	101	41.3	41.9	5.0	3.1	1.6	17.0	4.0	4.2	2.9	
12. Vision subn.	81	42.5	38.5	6.0	3.2	1.9	18.2	4.7	3.9	2.9	
13. Menstr. disord.	130	38.1	40.1	5.0	2.9	1.7	13.8	3.7	3.7	2.7	
14. Hemogl. below 70	62	37.4	39.0	0.0	3.6	0.0	13.1	4.4	2.8	1.4	*
15. Acne	140	36.1	40.9	—	1.7†	0.6	11.3	3.7	3.2		
16. Deformities	23	34.6	37.7	—	3.0	0.6	10.3	5.9	1.8		

*The average reliability of the difference is not given when Columns 5a and 6a give differences in opposite directions.

†The minus sign indicates that the average I.A. score for the item is lower than that of the group with which it is compared, and that the difference, therefore, lies counter to the "expected" direction.

‡1. The more severe, menstrual disturbances, necessitating going to bed during the period (B). 3. Hands: tremor, moisture. 4, 19. Students prevented, for health reasons, from participating in some forms of athletics. 5, 8, 12. For the criterion of muscular weakness, "poor posture" of students in these groups, the reader is referred to the original thesis. 9, 33. The groups include the students whose pulse rate takes longer than average to return to normal after a prescribed amount of exercise. 18, 32. Item 18 includes only the more seriously affected tonsils, item 32 all cases of defective tonsils. No such separation was possible in the case of women.

TABLE 2

Table showing the average I.A. scores of groups of men for whom a given physical characteristic or defect is recorded, and comparisons of these scores with scores of control groups.

Classification of difficulties recorded	Inferiority attitude score for group		S.D.	Comparison with the rest of the men <i>exclusive</i> of given group			Comparison with group of men for whom <i>no</i> positive findings are recorded			
	No.	Av.		Diff.	P.E. <i>diff.</i>	Diff. <i>diff.</i>	P.E. <i>diff.</i>	Diff. <i>diff.</i>	P.E. <i>diff.</i>	Average of 5c and 6c
Column number 1	2	3	4	5a	5b	5c	6a	6b	6c	7
17. Spinal curvature, poor pos.	372	36.2	42.2	7.8	2.3	3.4	10.2	3.5	2.9	3.1
18. Tonsils (operation advised)	35	43.0	42.9	10.2	5.0	2.1	17.7	5.8	3.1	2.5
19. Classification for athletics (B, C, D)	109	37.7	40.8	5.9	2.9	2.0	11.1	4.2	2.7	2.3
20. Knocked knees, bowed legs	83	37.5	44.3	5.5	3.5	1.6	11.5	4.6	2.5	2.0
21. Enlarged thyroid	88	36.6	43.0	4.5	3.4	1.3	10.6	4.4	2.5	1.9
22. Acne	141	35.4	41.9	3.4	2.6	1.3	9.4	4.0	2.3	1.8
23. Heart murmurs	26	38.5	37.9	6.5	4.6	1.4	12.5	5.9	2.1	1.7
24. Nasal obstruction	106	35.5	42.5	3.3	3.1	1.1	9.5	4.3	2.2	1.6
25. Phimosis	49	36.0	47.0	4.6	4.6	1.0	10.0	5.5	1.9	1.4
26. Defective hearing	111	35.8	40.9	2.2	3.7	0.7	8.6	4.2	1.9	1.3
27. Varicocele, hydrocele	88	33.0	36.9	1.2	2.7	0.5	7.6	4.2	1.8	0.9
28. Hernia	45	33.8	40.3	1.2	2.0	0.6	7.8	5.2	1.5	1.0
29. Deformities	35	32.7	44.5	0.3	10.3	0.03	6.7	5.1	1.3	0.6
30. Flat feet, fallen arches	47	33.3	41.3	0.9	4.2	0.2	7.3	5.1	1.4	0.8
31. Sugar or albumen in urine	50	30.4	39.6	—	1.7†	0.6	4.4	4.9	0.9	*
32. Tonsils (<i>all</i> positive records)	96	30.0	41.2	—	3.3	0.8	4.4	4.4	1.0	
33. Reaction to exercise	135	28.4	45.0	—	3.9	1.7	—	4.1	0.5	0.6
34. Vision subnormal	93	27.1	43.8	—	3.3	1.9	—	4.5	0.3	0.8
35. Undescended testicles	20	9.3	37.0	—	5.1	4.8	—	5.8	2.8	3.8

*The average reliability of the difference is not given when columns 5a and 6a give differences in opposite directions.

†The minus sign indicates that the average I.A. score for the item is lower than that of the group with which it is compared, and that the difference, therefore, lies counter to the "expected" direction.

TABLE 3
Table showing the average I.A. scores of groups of women who report
given difficulties (self-estimated), and the comparisons with control groups

Classification of difficulties recorded	Inferiory attitude score for group			Comparison with the rest of the women <i>ex- clusive</i> of given group			Comparison with group of women who report <i>none</i> of the difficulties			Average of 5c and 6c	
	No.	Av.	S.D.	Diff.	P.E. <i>diff.</i>	P.E. <i>diff.</i>	Diff.	P.E. <i>diff.</i>	P.E. <i>diff.</i>		
Column number	1	2	3	4	5a	5b	5c	6a	6b	6c	7
Worrying		157	60.6	36.1	33.0	2.4	14.6	36.4	3.4	10.5	12.5
Goiter		24	68.7	34.9	32.8	4.9	6.7	44.4	5.6	7.9	7.3
Insomnia		31	68.5	38.1	33.0	4.7	7.0	44.2	5.4	7.2	7.1
"Fair health"		25	65.3	33.2	29.4	4.6	6.4	41.0	5.2	7.8	7.1
"Tired"		136	53.5	41.7	21.6	2.7	8.0	19.2	3.7	5.2	6.6
"Nervousness"		38	56.8	35.2	20.8	4.1	5.5	32.5	4.8	6.8	6.2
"Palpit. of heart"		60	54.2	40.8	18.1	3.7	4.8	29.9	4.9	6.6	5.7
Gastro-intestinal disturbances		76	50.4	39.3	16.3	3.2	5.0	26.1	4.1	6.3	5.6
Headaches		54	53.6	39.9	18.6	3.9	4.8	29.3	4.7	6.2	5.5
Nasal obstruction		25	61.0	37.5	24.8	5.2	4.8	37.7	5.8	6.3	5.5
Poor appetite		31	58.2	41.9	22.1	5.2	4.3	33.7	5.8	5.8	5.0
Painful or frequent urination		16	52.5	34.9	15.6	4.2	3.7	28.2	4.9	5.8	4.7
Eye trouble		42	47.9	36.5	11.4	3.9	2.9	23.6	4.8	4.9	3.9
Frequent colds		123	43.2	40.2	7.8	2.8	2.8	18.9	4.0	4.7	3.7
Asthma or hayfever		22	46.8	39.0	9.8	5.2	1.9	22.4	5.7	3.9	2.9
Ear trouble		15	46.3	42.4	10.9	7.3	1.5	22.0	7.8	2.9	2.7
Skin trouble		32	42.3	43.0	5.0	5.3	0.9	18.0	5.8	3.1	2.0
Feet trouble		36	41.4	45.2	4.3	5.2	0.8	17.1	5.8	3.1	1.9

of inherited imperfection and susceptibility to disease which may remain latent, or again may break under the stress of circumstances. Every defect, no matter how trivial, is regarded by Adler as the expression and symptom of a general constitutional and hereditary weakness which focalizes in that organ. Moreover, it seems that among the specific defects Adler cites from the clinical literature many are no more serious than those discussed here.

When these physical items were tabulated for all the freshmen in the Liberal Arts College for whom inferiority attitude scores were also available, the following purely empirical method was resorted to. In the case of those conditions where the frequency of occurrence justified a statistical procedure, the average inferiority attitude scores were computed for individuals showing each condition, and the significance of the difference between this and the general average score determined. Thus the inferiority attitude score of women who were diagnosed as having an enlarged thyroid was compared with the average inferiority attitude score for the women who have no thyroid disturbance, and also with the average I.A. score of women for whom no defects of any kind were reported. These results are given in Tables 1-3.

This procedure is, of course, not free from objections. There is a very great amount of overlapping between the groups showing the various defects. That is, if the whole group is divided into those who have defective vision and those whose vision is normal, some of the other defects will be found in both the normal vision group and the subnormal vision group. Again, in comparing the defective vision group with those for whom no positive findings of any kind are reported, we must remember that the difference between the I.A. scores of these two groups may be a matter of general health as well as of the specific disturbance (defective vision) considered. The matter is further complicated by a purely medical consideration: we do not know which defects tend to appear together more frequently than others. Such an occurrence would produce, from the statistical point of view, the effect of correlated errors.

It is clear, then, that the absolute influence of any one defect cannot be isolated in this study. Only the relative amount of the influence can be approximated. In this case, as so frequently happens in statistical analysis of psychological data, we have to rely on numbers to indicate slight tendencies. We assume that if any factor, no matter how slight, is operative in a condition, and if all cases show-

ing this factor are chosen, some general tendency may be revealed, even though other factors tend to obscure the true state of affairs.

Three sets of data were treated in this fashion. Two of these are the defects diagnosed by the physician for each sex separately. These will be referred to as "objectively diagnosed" defects. The third set consists of women's "self-estimates" of their health.

So much for the analysis of the separate items of the physical examination records. But it would be of even greater interest to see if the general physical condition of the individual has any relation to his subjective feelings. Some general "health index" that would express health and vigor in a concise quantitative manner would be a convenient device to use; but no tool of the sort is available for the present. After the defects had been analyzed one by one, as described above, those for which the inferiority attitude averages were significantly high were chosen for further analysis. The data sheets were gone over, the number of such defects reported for each student was counted. These total numbers of defects reported for the individuals were used as a rough indication of the physical status of the individual at the time of the examination. Correlation coefficients were then computed between these indices and the I.A. scores.

Finally, before proceeding to the discussion of the results, a general statement must be made. The present study makes no attempt to interpret the physical examination findings from the strictly medical point of view. It recognizes that this should be done by men in the profession. All that the present analysis attempts to discover is whether or not any *very general tendencies* can be found to exist in the direction indicated by Adler.

RESULTS

A. Analysis of Separate Items

The results of the analysis of the separate items from the point of view of their relations to the inferiority attitudes are given in Tables 1-3. The results show that, for the most part, the items give statistically significant differences in the "expected" directions. That is, the average inferiority attitude scores for subjects with any given defect (the defects taken one at a time) are higher than the averages with which they are compared. Each set of data was analyzed in two ways; the average inferiority attitude score of all subjects showing a given defect was compared, *first*, with the average score of the rest of the group, i.e., all those *not* having this particular defect;

secondly, with the average score of those for whom *no* defects or positive findings of any sort were recorded.

Thus, in the list of objectively recorded items for women (Table 1), Column 1 gives the name of the item studied, e.g., menstrual disorders (B). Column 2 gives the number of cases in the group for whom this item is reported, in this case 119. Columns 3 and 4 give the average inferiority score together with the standard deviation of the distribution of scores for these 119 women. The following six columns give a comparison of the average score of these 119 with the rest of the group *exclusive* of those who have the given defect (5*a*, 5*b*, 5*c*), and with those for whom *no* positive findings of any sort are recorded (6*a*, 6*b*, 6*c*). Column 5*a* gives the obtained differences between the average score of these 119 women and the remaining 430 women—in this case the difference is 11.9. (A minus sign before the difference indicates that the difference lies counter to the “expected” direction.) Column 5*c* indicates the ratio of the difference to its probable error—in this case 4.0. This ratio is an indication of the significance or reliability of the difference. Columns 6*a*, 6*b*, and 6*c* are to be interpreted in a similar manner. In Column 6*a* the differences are spuriously high; here those who have flat feet, for example, are compared with those who have no defect of any kind recorded; but it must be remembered that it is the exception and not the rule to find on an individual’s card the record of flat feet and *no* other defects. Thus the exaggerated differences may be due *not* to flat feet alone, but to the cumulative effect of these other symptoms. For this reason, Column 5*c* probably presents the more representative values. But it will be noticed that the order of magnitude of the $\text{Diff.}/P.E.$ _{diff.} in the two columns remains approximately the same. For this reason it would seem that it makes very little difference which criterion of comparison is used.

The data for the men and the women are not, strictly speaking, directly comparable, but some general statements can be made about both. In both cases the items relating to general physical strength—as strength of muscles, posture, gymnasium classifications—have about the same average $\text{Diff.}/P.E.$ _{diff.}, and especially for men come rather high in the series. The “eye, nose, throat” defects also have similar differentiating values within both groups, and are far down on the lists for both men’s and women’s defects. Although it has been possible to give a general characterization of these two groups of defects—those relating to physical vigor and to “eye, nose, and

throat"—it would seem from this particular set of data that no one physical symptom has any striking relation to the inferiority attitude.

It is to be noted that in the case of men the items related to the urino-genital apparatus have on the whole low indices of reliability of the difference. They are as follows: 1.0, .6, .5, and -4.8 (5c). The last item, which shows a difference that runs counter to the "expected" direction, refers to the average inferiority score of men who have a record of "undescended testicles." Their average inferiority score is 9.3, which is over half a standard deviation below the average of the whole group. Yet these are the very characteristics which, according to the literature, we would expect to give rise to feelings of inferiority and to show significant differences. Here again we encounter the difficulty of having to differentiate between two possibilities. On the one hand, we may have a direct relation of the general constitutional inferiority as described by Adler to a general feeling of inadequacy because of lack of energy and vitality; on the other hand, we may have a sensitiveness about a particular physical defect about which feelings of inferiority might be focalized. These two possibilities are, of course, not mutually exclusive.

It is true that most of the items reported above are not of frequent occurrence; in particular, the item "undescended testicles" is reported for only 20 men. But if a very pronounced relationship existed we might expect some trace of it to be apparent in even a small group. It may be that the defects related to the urino-genital apparatus are of no greater importance (qualitatively or quantitatively) than any of the other physical characteristics. It may be, finally, that in these cases some over-compensatory reactions are developed which are not adequately measured by the scale.

It has been mentioned before that the data for men and for women are not directly comparable. Table 4, which shows the percentage frequency of occurrence of the defects in the sexes separately, shows in some cases differences in the percentage of occurrence which are more likely due to the dissimilarity of standards used than to some difference in physical condition. Thus "nasal obstruction" is reported as occurring in 26% of the women, but only 15% of the men. Fallen arches and flat feet occur in 14% of the women, but in only 7% of the men. Heart murmurs are reported for 15% of the women and 4% of the men. Percentages are, in general, higher for the women; this may be due to more careful medical examinations.

TABLE 4

TABLE SHOWING THE PERCENTAGE FREQUENCY OF OCCURRENCE OF DEFECTS
IN THE SEXES TAKEN SEPARATELY

Women (total-549) Classification	%	Men (total-680) Classification	%
Fallen arches and flat feet	14	Flat feet	7
Heart murmurs	15	Heart murmurs	4
Orthopedic gymnasium	19	Classification B, C, and D	16
Enlarged thyroid	32	Enlarged thyroid	13
Poor posture	38	Poor posture	55
Reaction to exercise	13	Reaction to exercise	20
Nasal obstruction	26	Nasal obstruction	15
Tonsils	18	Tonsils (all cases)	14
Vision, subnormal	14	Vision, subnormal	13
Deformities	4	Deformities	5
Acne	25	Acne	21
Hands, tremor and moisture	29	Knocked knees, bow legs	12
Menstrual disturbances (B)	22	Defective hearing	16
Menstrual disturbances	24	Albumen and sugar in urine	7
Hemoglobin, 70 and below	11	Hernias	7
Strength of muscles	20	Varicocele	13
		Phimosi	7
		Undescended testicles	3

Thus the closer relationship of the defects to inferiority in the case of women (as indicated by the higher *Diff./P.E.*_{diff.}'s) may be due in part to the greater accuracy with which the data were collected.

An interesting problem is raised when the two sets of data on women—the defects recorded by the physician and those reported by the students themselves—are compared.

On the whole, the self-estimates show greater and more statistically reliable differences. The comparison of the average diagnostic values of difficulties which were recorded in the two ways is given below.

Even here, except for nasal obstruction, the difficulties refer to

TABLE 5

Physician's diagnosis	<i>Diff./P.E.</i> <i>diff.</i>	Self-estimate	<i>Diff./P.E.</i> <i>diff.</i>
Acne—skin trouble	0.6	Skin trouble	0.9
Vision	1.9	Eye trouble	2.9
Heart murmurs	3.2	Palpitation of the heart	4.8
Enlarged thyroid	2.8	Goiter	6.7
Nasal obstruction	2.2	Nasal obstruction	4.8

slightly different things—but the fact remains that the indices for the self-estimates are by far the higher of the two. This might be brought about by one of two things: (*a*) either the students who report these difficulties themselves are the ones who possess the symptoms in a more severe degree and are thus made aware of them, (*b*) or it may be that the severity of the symptom is the same in both cases, but that those students who estimate their difficulties as serious enough to report them are the more unstable individuals whose I.A. scores are higher, who take these minor ailments more seriously and who are more concerned about their own health. The true state of affairs cannot be made clear from the data reported here. To analyze this complex relationship, it would be necessary to have access to very carefully made medical diagnoses where the severity of the symptoms, as well as their mere presence, was scrupulously recorded. The distinction is made here simply as an indication of an interesting problem related to hypochondria, which needs further study.

On the whole, the uniformly low numerical values of the differences and their reliabilities suggest the following explanations: that it is the *general* condition of being in *any* way below par physically rather than the possession of any *particular* defect that shows a relation to the inferiority attitude, that the *nature* of a defect is not important in determining inferiority attitudes; that this attitude may be associated with one kind of defect as well as another; and that a defect may be present without being accompanied by the feeling of inferiority. This evidence, along with other data on the attitude measured (4), suggests the principle of multiple causation in the interpretation of the inferiority attitude.

B. Total "Defect Scores" and Inferiority

In the case of the women the total defect scores were obtained in several ways: (*a*) by counting as defects all positive findings except those which, according to the doctors, were of no significance (this was done for the results of the medical examinations only); (*b*) by counting as defects, for each individual, all the items which gave a *Diff./P.E.*_{diff.} of 2 or more according to each of the two criteria used.⁵ This was done for both the physician's records and for the

⁵The criteria of comparison refer to comparisons with students for whom no defects of any kind are reported, and comparisons with the rest of the group exclusive of the given defect. See Tables 1-3 and their interpretations.

self-estimates, making a total of two "objective" scores and two "subjective" scores for each individual.

In the case of men, only objective records were available. Since the *Diff./P.E.*_{diff.}'s are all very low (Table 2, Columns 5c, 6c), and no more than three indices in any column exceed 2, the total defect scores for the men were obtained in only one way: by counting *all* the items except those which the physicians consulted pronounced as of no importance. Correlations were then computed between "defect scores" obtained in each of the ways described and the inferiority score. Since the appearance of the scatter diagrams suggested the possibility of a non-linear relationship, both the Pearsonian *r*'s and the correlation ratios were computed. The results are given in Table 6.

TABLE 6
CORRELATIONS BETWEEN TOTAL NUMBER OF PHYSICAL DEFECTS AND THE INFERIORITY ATTITUDE

Women—Objectively Recorded Items.* N=549		
	According to Column 5c (Table 1)	According to Column 6c
<i>r</i>	.22±.03	.21±.03
eta	.24	.22
Women—Subjective Estimates.* N=549		
	According to Column 5c (Table 2)	According to Column 6c
<i>r</i>	.29±.03	.30±.03
eta	.34	.34
Women—Items Estimated as Significant by the Physician		
<i>r</i>	.24±.03	
eta	.27	
Men—N=680		
<i>r</i>	.16±.03	
eta	.17	

*Including items for which the reliability of the difference is at least 2.

Inspection of the tables of correlation coefficients shows that when the data are analyzed in this gross quantitative manner, the relationship persists, slight but very definite. Considering the number of cases on which the correlations are based—680 men and 549 women—it can be said that correlations of .16, .22, and .29 definitely establish a general tendency in the direction indicated.

Examining these results in somewhat greater detail, we notice first the sex difference which has appeared all along. The relationship is more definite for women than for men. It is to be expected, of course, that when this difference appeared for the separate items, it would also persist for the items considered together.

In the second place, in the case of the women, the correlations for the "self-estimates" are higher than the correlations for the objective records, although the difference is not large. This again is to be expected since the separate "self-estimated" ailments gave higher *Diff./P.E.*_{diff.}'s than the objectively recorded ones.

On the whole, the correlations are surprisingly stable; no matter how the items are chosen, the value of the coefficients remains approximately the same. This further supports the hypothesis that no *one* physical characteristic bears a close or necessary relationship to the inferiority attitude.

As to the character of the relationship, there is no evidence that it is non-linear. It is true that in most cases the correlation ratio gives a slightly higher value than the Pearsonian r , showing that the curvilinear regression line might fit the data more closely. But in only one of the cases is the difference between the Pearsonian r and the correlation ratio statistically significant. The appearance of non-linearity, which suggested the advisability of calculating the correlation ratios, may have been due to the fact that the "defect scores" were slightly skewed toward the lower end of the distribution—the *modes* fell at 2 while the total ranges were from 0 to 7 or 8 defects.

In both the objectively and subjectively recorded items, the inclusion of a larger number of items does not appreciably alter the correlations.

The correlations of inferiority scores with the total number of items considered significant by the physicians is not very different from the correlations based on items chosen from the lists on the basis of the statistically and empirically determined relationships to the inferiority attitude.

SUMMARY

1. When the positive findings recorded on the medical examination blanks are analyzed from the point of view of their relation to the inferiority attitude, the majority show differences in the "expected" direction; that is, the individuals for whom any given item is recorded show higher average inferiority attitude scores than those of the unselected freshman class.

2. The relationship holds when the "total defect scores" are correlated with inferiority. The correlations are .16 for men and between .20 and .30 for women.

3. Sex differences exist in the sense that the relationship is in

every case less clear for men than for women. This may be due in part to a difference in the way in which the medical data were collected.

4. In the case of women, the "self-estimates" gave a slightly higher relationship than the items recorded by the physicians.

These results indicate, therefore, that the relationship between organic inferiority and feelings of inferiority, which Adler finds in his clinical studies, is also found to exist, as a slight but consistent and definite tendency, in the "normal" population, when the data are analyzed in the manner described here.

REFERENCES

1. ADLER, A. Study of organ inferiority and its psychical compensations. (Trans. by S. E. Jelliffe.) *Nerv. & Ment. Dis. Monog. Ser.*, 1907, No. 24. Pp. 86.
2. ———. The neurotic constitution. (Trans. by Bernard Glueck and John E. Lind.) New York: Moffat, Yard, 1917. Pp. 456.
3. ———. The practice and theory of individual psychology. (Trans. by P. Radin.) New York: Harcourt, Brace, 1924. Pp. 352.
4. FATERSON, H. F. A study of the Minnesota rating scale for measuring inferiority attitudes. *J. Soc. Psychol.*, 1930, 1, 463-493.
5. HEIDBREDER, E. F. The normal inferiority complex. *J. Abn. & Soc. Psychol.*, 1927, 22, 243-258.

Smith College
Northampton, Massachusetts

L'INFÉRIORITÉ ORGANIQUE ET L'ATTITUDE D'INFÉRIORITÉ

(Résumé)

On a fait une analyse de l'Echelle d'Evaluation Minnesota pour la Mesure des Attitudes d'Infériorité en rapport des résultats des examens physiques des étudiants universitaires de première année. On a trouvé l'existence d'une relation positive très définie quand on analyse les caractéristiques physiques spécifiques séparément, et quand on fait la corrélation des résultats de "défaut total" avec les résultats de l'Echelle d'Evaluation. Quand on analyse les défauts et les caractéristiques rapportés par les étudiants eux-mêmes, la relation devient beaucoup plus marqué que quand on analyse ainsi les résultats objectifs, c'est-à-dire, les diagnostics faits par les médecins. La tendance générale des résultats indique que la relation entre l'infériorité organique et les sentiments de l'infériorité trouvés par Adler dans ses études cliniques se trouve aussi dans la population "normale" quand on analyse les données de la manière décrite.

FATERSON

ORGANISCHE MINDERWERTIGKEIT UND MINDERWERTIGKEITSVERHALTEN

(Referat)

Es wurde eine Analyse der Minnesota Messungsskala (Rating Scale) zur Bestimmung der Minderwertigshaltung vorgenommen und mit den Ergebnissen der physischen Untersuchung von Erstjahr-Universitätsstudenten verglichen. Es ergibt sich eine geringe, immerhin sehr bestimmte positive Beziehung zwischen beiden, wenn die spezifischen physischen Merkmale getrennt untersucht, und wenn die "Gesamtminderwertigkeitsquotienten" (total defect scores) mit den Ergebnissen der Messungsskala verglichen werden. Wenn die von den Studenten selber angezeigten Defekte und Merkmale untersucht wurden, war die Wechselbeziehung viel ausgesprochener, als wenn objektive Angaben, dh. ärztliche Diagnosen untersucht wurden. Die allgemeine Richtung der Ergebnisse zeigt, dass die Beziehung zwischen organischer Minderwertigkeit und Minderwertigkeitsgefühl, die Adler in seinen klinischen Untersuchungen findet, auch in der "normalen" Bevölkerung existiert, wenn die Angaben in der oben beschriebenen Weise analysiert werden.

FATERSON

SHORT ARTICLES AND NOTES

PSYCHOLOGICAL REACTIONS OF BIOLOGISTS: A CROSS-SECTION OF PRESENT-DAY BIOLOGICAL THOUGHT¹

EDWARD J. V.K. MENGE

In an investigation of the more important problems of the day as the biologist sees them, and in a study of the vital facts of biology as the biologist himself evaluates them, there is practical unanimity. In attempting to read meanings and interpretations into the facts of biology, however, one meets with considerable differences among even the ablest biological workers. Consider the case of the Hertwig brothers, Richard and Oscar, internationally famous biologists, whose differences in the matter of interpretations has so well been presented by one of our younger biologists, Dr. John Giesen, in his *Backgrounds of Biology*. Here it is shown that these two brothers, children of the same parents, having the same environment and the same training, going to the same university, majoring in the same subject, and having the same professors, nevertheless, came to quite opposite conclusions from the very same facts. Richard became an ardent defender of Darwinism and Oscar became an intrepid foe of that same doctrine. Even after a half century in continuous biological work, Richard was still insistent that Darwinism, adopted in his youth, was the great basic principle that gave the best explanation of how things came to be what they are, while Oscar wrote a most uncompromising attack on that same doctrine.

If the ablest of men, with the same inheritance, the same environment, and the same training, who have made a lifetime study of a subject, cannot agree, what hope is there for others, who know little or nothing about a subject, to evaluate the facts validly?

It is not strange, then, that in attempting to collect and evaluate the opinions of things biological as the foremost biologists of the world interpret them that one should run the entire gamut of possible explanations, from that of the writer who flatly and with decided finality insists that biology, as such, never has, does not now, and never can influence anything or anybody, to the one who insists with the most enthusiastic avowal that every advance the world has made, is making, or ever can make, must always, in the final analysis, rest upon the findings of biology and biologists.

¹This paper furnishes a summary of what the ablest men in the biological world hold at the present moment. It has been arranged from not only the general literature but also from a series of letters written the author by the heads of departments of biology and zoology of the leading universities of the world. It represents the views of more than two hundred correspondents.

On the positive side we find, from our survey of biological opinion, that the most outstanding and important findings of biology which have influenced mankind are summarized under the following headings:

1. Medicine and general health, especially through the study of parasitology and disease prevention. In this connection, one must not forget the remarkable results obtained in the practical fields of medicine from such studies as those made in malaria, yellow fever, typhoid fever, and diphtheria, to mention only a few.

2. Bacteriological and parasitological findings applied to commercial processes, and insects and plant control, which, in turn, affect the entire food supply of man.

Closely connected with this latter work is that of applied genetics. Here questions like the following find their answers:

"How can we breed those types of plants and animals which will thrive best and present those characteristics we desire in our own particular region of the world?"

"How can we make disease-resistant plants and animals?"

"How can we make the waste places of the earth become of value to man?"

"How can biology be brought to advance the work of the agriculturist, the cattle breeder, etc.?"

"How increase the yield of crops?"

Thinking of matters of this kind, Professor Gardiner says:

"Our knowledge of vitamins and their production has undoubtedly changed the lives of people in this country (England) very profoundly; its association with the knowledge of feeding and growth in animals and plants has influenced the development of agriculture on every side."

The effects "are reflected in politics and education. Scientific biology is at the base, and stimulated the development, of fisheries not only in Western Europe but in your own country (the United States) as well; the basal work and the stimulators were university men of science, not people paid by the state."

3. The beginning of a so-called human biology in which the developments and findings of comparative psychology assist us toward a better understanding of human relations.

In this regard, Professor G. H. Parker holds that

"Modern aspects of animal activities in relation with genetics will serve as a basis for understanding human life in a way in which it has never been so completely understood heretofore. The biological attack on this question is probably the most fundamental that has ever been made."

Here one must consult the application of these findings to sociology at large. One of the outstanding questions which confronts us and requires an answer which the biological psychologist alone can give is: "What and why were certain variations and modifications allowed to persist which could not have persisted without some social shield?" Here, too, must be placed those newer findings and interpretations resulting from the study of endocrines and their relation to the emotions. This is a second fundamental

problem of human biology, and one which as yet is scarcely understood, but nevertheless is bound to influence most profoundly much of our present and future interpretation of all things psychological and sociological.

4. Organic evolution; a subject mentioned by practically every writer. Professor B. A. Bensley says in this regard:

"I do not believe that the larger effects of biological or other scientific study can be of greater or different consequence in one country than another. I think we would have to admit that organic evolution and its historical setting have made not only the greatest difference to humanity, but also that no single issue otherwise, including the entire industrial development of the world, has been so important. Organic evolution ought to have been one of the minor issues—historically the stage was not set that way."

5. To the four headings mentioned above, Professor Conklin adds a fifth, the non-inheritance of acquired characteristics, a biological factor with a vast influence on account of its wide ramifications in almost every walk of life.

On the negative side, we may quote one correspondent who reviews the matter thus:

"To my mind the greatest achievement of biology, which is at last being realized, is negative: that science, when placed in any sort of straight-jacket (of preconceived theory), leads into blind alleys."

This fact is instanced in the following cases:

a. In embryology, the mistake of interpreting all phenomena of development as ancestral indications is now generally admitted. All this is well shown by the writings of such men as T. H. Morgan, Wheeler, Jennings, Ritter, and a number of others.

b. In genetics, the work of Jennings, Conklin, Morgan, and others has shown that environment and heredity stand on an equal basis. Heredity is not the all-powerful influence formerly taught, and environment nothing. Both are coequal.

c. Physiology cannot achieve results by ignoring anatomy. A structure functions only in relation to an environment (internal or external). Physiology, by disregarding anatomy and ecology, has proceeded no whit farther from the position Goethe criticized in Faust a hundred years ago with the words,

"Wer will was lebendiges beschreiben
Sucht erst den Geist heraus zu treiben.
Encheiresin naturae nennt die Chemie
Spottet ihrer selbst und weiss nicht wie."

Organisms cannot be understood by studying the anatomy, the environment, or the functions alone. There must be interdependent study. The same writers mentioned above also have written on this matter.

4. "Mechanistic evolution" is a straight-jacket for experimental work. "Emergent evolution" permits mental freedom. However, if the writers on this theme permit themselves to go on logically, it will drive them di-

rectly into the philosophical position of the neoscholastics. Evolution is not an automatic thing, but an intelligently guided phenomenon. One thinks of the names of Lloyd Morgan, Ritter, Lovejoy, and Noble, especially, in this regard.

5. Too much time, beginning with Darwinism, has been wasted in useless speculation. Even now, too many biologists are spending time in speculating, instead of applying the experimental facts to their pet theories. Biological philosophy must be fitted to the facts, not facts to preconceived notions. This, too, is just beginning to be realized by such men as Jennings, Morgan, Wheeler, and a host of others.

6. The line of human descent leads away from the apes, rather than to them. That our primitive ancestors had a *mind*, hence intelligence, is revealed by additional human discoveries; wherever man has left traces besides his own bones, those traces show the workings of an intelligence. These same traces show that the "cave man" of popular fancy never existed: the cave man was an extraordinarily intelligent man, endowed with a mind: one who had a religion, believed in a hereafter, honored his dead, had a knowledge of handicrafts, and who practiced various arts (painting, sculpturing, modelling, carving, etc.). Osborn recently remarked in this regard, "Man is the most incomprehensible thing of this world; and in man the mind is the most incomprehensible thing of all. There is no transition between him and other animals; even his so-called nearest relatives show not the slightest trace of intelligence."

The intellectual man really seeks philosophical progress. As one correspondent puts it:

"Philosophical progress is the nearest thing to what our nature seems to demand—perhaps one might say an up-to-date religion. I would say that all demonstrations, especially geological, palaeontological or in general, truly historical that will yield facts rather than sympathetic theories, will be important but accessory. Some are tired of evolutionary work, others think the picture is not plain enough to satisfy.

"I would, therefore, place in order of importance, first the demonstration of chemical homologies and sequences in living protoplasm. I believe they will provide a counterpart and largely an explanation of the ramifications of morphological modifications. Second, the primary nature of living protoplasm. These two are related, but I think there is no doubt that the sticking point in biological philosophy is the origin of life."

We find the biologist's spirit depressed by the fact that the intellectual phases of scientific progress enter so slowly into the actual life of humanity. Whatever influence biologists may wield on the political, social, and even religious thought of the day seems to come through some more or less practical applications, as only that very small, more highly educated portion of the public ever can approach a true conception of the scientific philosophy of the time.

As Professor McClung (6) wistfully puts it:

"At a time when science is proclaimed the chief reliance of organized society in securing perpetuation, when through its ministrations, human life is materially lengthened, made more effective and enjoyable: when the uncertainties of existence, and the terrors of the unknown are yearly being reduced in significance, then we witness the paradox of vicious and unreasoning assaults upon the methods and conclusions of science by legislative enactments to cripple its progress and limit its teaching.

"It would seem from all this that science is in our day and generation but not of it. This is no doubt largely the fault of scientists who have ever been inclined to become absorbed in their pursuit of knowledge and to manifest little concern with the use that is to be made of it."

Certainly nothing is added to the intellectual probity and honesty of scientific men, when some of our most important publishing houses are printing editions of biological texts in which every reference to evolution is deleted, just to please the less intelligent regions of the country. Such a procedure is especially bad when everyone knows that both author and publisher believe things quite different from what their textbooks imply. If scientists will cast aside what they really believe and are supposed to hold dear, merely to sell a few extra copies of their publications, why should one expect non-scientific minds to believe that science leads to truth?

Peabody (7) has recently said:

"The popular conception of a scientist as a man who works in a laboratory and who uses instruments of precision is as inaccurate as it is superficial, for a scientist is known not by his technical processes but by his intellectual processes; and the essence of the scientific method of thought is that it proceeds in an orderly manner toward the establishment of a truth."

Yet how far short of that ideal even a most eminent man in such a supposedly absolute science as mathematics may fall! Consider the case of M. Michel Chasles (8, pp. 10-11), one of the leading geometers of his time. Between the years of 1861 and 1870 he bought more than 27,000 manuscripts forged by a man with meager education, who must have worked many hours daily for long years to complete so many. Here were hundreds of letters from Pascal, twenty-seven from Shakespeare, hundreds from Rabelais, many from Newton, Louis XIV, from the Cid, and from Galileo. There were letters from Sappho, Virgil, Caesar, St. Luke, Plato, Pliny, Alexander the Great, and Pompey—even a letter from Cleopatra to Caesar, a note from Lazarus to Peter and a pleasant little missive from Mary Magdalene to the King of the Burgundians.

Every one of these letters was written on the same kind of paper (not on parchment), and all were written in French. One can appreciate a non-scientific writer's comment upon this case,

"Remember that these manuscripts were eagerly purchased and their authenticity warmly defended by one of the leading geometers of his time, and then believe, if you can, that development of the mathematical faculty has anything to do with reasoning power, or even with common sense."

When we speak of men still living, or but recently passed away, in whose disciples the fires of emotional desire for credit looms large, matters of importance to the real advancement of science are all too often confused with a mere matter of priority, so that, as Metcalf has pointed out, the more valuable work is usually overlooked or forgotten.

To this same idea J. M. Aldrich (1) contributes the following:

"Criticism is usually a generation behind the publication, so that the poorest work may stand, as we often see, for twenty or forty years before it is completely revised." Moreover, in North America at least, the ablest men are quite likely "to be the very ones to succumb to the higher rewards of administrative and economic work, and so fail to make the contribution of which they are capable."

And finally, in the matter of psychological reactions of scientists, Poincare has called attention to the comparison of Gallic and Anglo-Saxon genius:

"Characteristic of the one is a feeling for form, for symmetry, for logical completeness, for finality; characteristic of the other is the feeling for substance, development, function, change. For the one, truth lies in the result; for the other in the process. One is represented by the deductive, the other by the inductive type of mind."

The one type of mind holds that there is a possibility of finding an exact correspondence between the thing itself and our idea of that thing, which means absolute truth insofar as that one idea is concerned, at least. The other holds that there is no such thing as absolute truth or falsehood of any theory or observation but that, as scientists, it is our duty to weigh carefully all the chances for truth and error, place each in its proper position, and thus improve the way. Professor Beach (2), speaking for this group, puts it like this:

"It is important to recognize the fact that scientific laws are not proved by perfect corroboration of measurements. The proof of any law is of negative character. Not even the law of gravitation, nor the law of the conservation of energy is proved by positive demonstration. The probable truth of any proposition is assumed from the inability to disprove it. Whence it follows that there is nothing more fundamental to the correct understanding of the science of physics, or indeed of science in general, than the interpretation of measurements according to the theory of probabilities and a rational discussion of the inherent errors."

H. B. Torrey (10), speaking of the influence that modern scientific thought has had on philosophy, adds:

"Science has progressed in the biological sciences by abandoning a faith in final causes for a faith in the hypothesis that works, by draining off every stagnant suspicion of ultimateness in explanation, in the light of the conviction—the product of experience—that the ideas that serve us change with our knowledge of object facts."

And I think that he sums up well the ideas on this matter of the average biologist in the three following paragraphs:

"For one who seeks a basis of criticism for a contribution to science, three obvious tests may be applied.

- (1) It may contribute new facts.
- (2) It may contribute a formulation of old facts.
- (3) It may contribute a new idea that, in the presence of facts, may lead to a new point of departure for explorations into the unknown.

"If, then, one were to apply these tests to what seem to me to be the two most significant developments in the philosophic thought of today, they might be said to fall, very roughly speaking, under the second and third categories. In the former, might be placed the synthetic philosophy of Spencer, an avowedly scientific philosophy whose essential problem was to formulate the known facts of science in terms of principles of evolution. This stupendous project, remarkable alike for the powers of its author and the wide range of his interests, ended in a system of philosophy, into which just enough metaphysics succeeded in creeping to justify the criticism that, in spite of all good intentions, he had not been able to disentangle himself completely from the habits of thought to which his critics were happily accustomed.

"Pragmatism distinguishes itself at once from the synthetic philosophy in that it is nonsystematic. Instead of an interest in a formulated body of knowledge it appears to possess an insatiable desire to determine practical choices. *Given a problem of conduct, the solution unknown; what shall the line of action be? Here one perceives a strictly scientific situation that emphasizes the practical value of the hypothesis. The problem is to find a satisfactory path into a new region. And the answer that pragmatism gives is, trust to luck and your past experience.* 'The truth,' says James, 'is the hypothesis that will work.' 'The truth' says Dewey, if I rightly apprehend him, 'is the hypothesis that you can work with.' There is a suggestion of permanency, of stability, of future significance in the latter phrase that makes it, to my mind, more felicitous. But I do not care to dwell on that point. What comes closer to my purpose is to point out that here is no faith in final causes, here is no suspicion even of that innocuous phantom, the unknowable. Here is no distinction between science and philosophy—if indeed pragmatists are philosophers, in spite of the fact that, in one form or another, they fill several chairs of philosophy now in our universities. Here is a faith that facts will tell their tale—will inevitably condition the movement of ideas, that one's imagination content is derivable from one's effective experience. Here is a philosophy that is working a transformation on the thought of the day. How? By abandoning the search for lofty peaks of final causation, from which to triangulate the universe according to logical necessity; but emphasizing ideas that shall not only square with the facts as we find them, but shall create others."

Although this may be the attitude of the average biological worker, there is another group, none the less able, whose logic leads them another way.

I cannot help mentioning here a conversation with a scientific surgeon held a year or two ago. A matter of surgical procedure was being discussed. The surgeon admitted the force of his opponent's argument. In fact, he said, "There isn't a flaw in the reasoning. I accept the logic of it, and yet, if this happened to one near and dear to me, I would not follow the dictates of my logic merely to make a logical conclusion come true."

To me it seems that here is the crux of a vast amount of conflict in the scientific world. One man will follow consistently wherever his logic leads him, feeling that he cannot do otherwise; the other feels that every problem is to be judged entirely by itself, without relation to the larger issues of

history, philosophy, psychology, logic—in fact, without regard to all that which has made our entire background of education and training what it is.

Professor John Dewey (4), in sketching the development of the philosophies in various countries, laid heavy stress upon the fact that the United States of America had not given sufficient thought to philosophy. He adds that, "as long as we worship science, and are afraid of philosophy, we shall have no great science except a lagging and halting continuation of what is thought and said elsewhere."

We may all be wrong in our interpretations. Let us not forget that in the time of Columbus the ablest men were certain that Columbus was mistaken. A later generation may feel the same toward us. The best we can say in all truthfulness is that while we may be mistaken in everything we now hold dear, yet one is more likely to be correct more often by following the abler men than one is in following those less able. With this as an interpretative background, let us continue with a report of our findings.

The final question in our discussion is that of the status of the vitalistic and the mechanistic conceptions of life. What do the ablest biologists in the various lands hold regarding these two ways of explaining living organisms?

Before presenting their answers, let us suggest precaution to the non-philosophical reader. A vote can never settle anything as to the actual facts in any given case. All that it can do is to show what the majority hold.

We have seen how proximity to one's own work and to those who think alike causes an exaggeration of emphasis on such work and such thought. Scientists, like other people, are so close to their own work that they are like the man who could not see the forest for the trees. A wider outlook, it is true, is discernible in the writings of many who are trying to treat the organism as a whole and not as a mere collection of individual and distinct parts.

Thus, Vernon L. Kellogg (5) holds that, no matter how the scientist may look at a thing from a purely physico-mechanical point of view, when he sits back in his chair in his own home and views his wife and children, somehow there is a greater gap between their reactions to him and his to them than his laboratory findings would lead him to believe.

Professor Kellogg thus holds with Balzac that, regardless of what the laboratory tells us, humanity will never accept the implied ultimatum of the alchemist who says to his weeping wife, "Stay, I have decomposed tears. Tears contain a little phosphate of lime, some chloride of soda, some mucous and some water." Is there nothing more to a tear than this? Will humanity ever be willing to accept that analysis? In all probability it will not. Unless the findings of the laboratory fit in with the findings of humanity as a whole, it will do little good to attempt to force acceptance of them with any degree of finality.

If, as Wm. E. Ritter (9) says, the golden rule has influenced the actions of mankind, it is mighty poor science which would neglect it in writing the history of mankind. It would be like studying the bee and forgetting to mention one or more of the important traits found. And so, T. D. A. Cockerell (3) very recently laid emphasis upon the fact that mechanistic science can never bring out a sense of human values. One can, of course, always say that science shows that the race must be protected, and that consequently one can work out some scheme of morality and a sense of duty from that angle—that what is best for the race and what is best for posterity should be done. But, doesn't such a scheme only furnish a *motive without any obligation*? Suppose I care not a whit for the race or for posterity, then why should I hold it a duty to sacrifice even a little convenience for them?

Mankind vaguely reaches out for something positive in the way of an analysis of *duty*. Nikolai recognized this in his *Biology of War*. That entire volume was written practically just to make the point, that something must furnish obligation if there is to be any *duty*. And yet, astonishing as it may seem, after going over *every* review of this book, in every journal written in English (and there were many reviews), I cannot find a single reviewer even so much as mentioning this very thing about and for which the book was written.

We seem to find that scientists are of two types. The one insistent upon remaining quite modest and laying emphasis upon how little we know about anything; the other, not so modest, who insists upon laying all the emphasis on the great progress that has been made.

Even when we dissect Poincaré's division of types of mind into the Gallic and the Anglo-Saxon,² we find the philosopher Schwegler insisting that the Anglo-Saxon mind is extremely acute but seldom profound.

²Probably nowhere can the difference in the temperament of two peoples be shown better than in group reactions of the Anglo Saxons and the Latin as shown in the following popular account of understanding and lack of understanding of each other by Duff Cooper, M.P., in a recent issue of *Graphic*, London:

"The difference between French mobs and English mobs, and between political disputes in the two countries was never more vividly exemplified than during the general strike in England. For ten days England seemed to be on the brink of civil war—the eyes of the world were turned toward us in the expectation of terrible developments—and when it all ended foreigners could hardly believe that not a single casualty had occurred.

"While it was still in progress, there occurred in Paris some celebration in connection with their national saint, Joan of Arc, an historical figure for whom it might be supposed that even an anarchist could feel nothing but respect. But riots broke out, and in one afternoon loss of life and many casualties occurred.

"Meanwhile the strikers in this country were playing football with the policemen, a phenomenon which so puzzled and therefore exasperated foreigners that they were driven to explaining it by saying that the English were incapable of taking even their own revolutions seriously."

He holds that this latter type of mind can find more and devious ways down which to travel than any other group, but seldom can it stay at one problem until such problem is plumbed to its very depth.

It matters not whether we accept any of these analyses. It does matter whether or not we know that they are there, and that the men who hold them are just as able as those who hold opinions quite the reverse.

The one type of mind insists upon knowing everything about a subject; the other wants to know the subject itself. Let us illustrate. It is conceivable that one might find a scholar who knew every book and article ever written on friendship; who had taken those who were friends of each other into his psychological laboratory and found how they reacted to each other under all and sundry conditions. In other words, here is a man who knows more than anyone else in the world on what we may call the *scholarship of friendship*, and yet, it is also conceivable that he himself never has known *friendship*. Knowing *about* a thing, and knowing *it*, are two different things.

The poet who has sung of

"A sense of law and beauty, and a face turned from the clod,
Some call it Evolution and others call it God,"

has really a wider understanding of that thing we call life than has many a biologist. For, after all, has he not in these few lines really described life? The law of which he speaks is science, and the beauty he mentions calls forth the idea of ideals, of character, in fact, of all things which make us appreciate life to the fullest, whenever and wherever it may be found.

Of these outstanding biologists from whom I have sought opinions, I find but comparatively few who are willing to go on record as out-and-out mechanists. Most of them probably feel as the educated people of the United States feel when prohibition is discussed—that no one with any intelligence can be a dry and no one with any decency can be a wet. So many crimes have been committed in the name of each that few are willing to align themselves as partisans. Thus it seems to be with Vitalism and Mechanism.

Probably the best summary of those leaning toward the mechanistic side is shown by the statement of one who said "most biological workers are *practical mechanists*, though they admit that life may be something more than mechanism." Or, by another, phrasing it thus: "they are mechanistic in practice but agnostic in opinion." Another adds, "It is always dangerous to make categorical classifications, but a majority of American biologists are not disposed to believe in Vitalism, if by that is meant that there is in living matter some principle entirely different from that found in the organic world, that is to say, some principle entirely outside the range of physics and chemistry; on the other hand, I do not know that these workers would care to be styled mechanists, because of the rather crude ideas that have come to be associated with the term."

But the opinion of men who stand at the very heights in biology, in America and Germany at least, may be summed up in this statement:

"Vitalism and Mechanism seem to be antiquated ways of looking at the problem of organic activity and will be replaced by more modern views on this subject. The contrast between Vitalism and Mechanism is probably a fictitious one. An irresistible trend of thought and feeling in biology based on the aggregate of positive results gained in the various subdivisions of technical research will, I am quite sure, bring about the final disappearance of the old conflict between Vitalism and Mechanism with no out-and-out victory for either side.

"This conflict has been due largely to the prevalence in it of a maximum of speculation and a minimum of common sense. That many of the structures and activities of any individual animal are mechanical in the sense that they act to a large extent according to the principles of mechanics is so obvious that nobody questions it as long as he thinks and speaks as an ordinary mortal. Compare the workings of our own limbs with the working of levers, for instance.

"But even the hard-boiled mechanist never questions (so long as he remains true to himself just as an every-day human being) that the individuals in which these mechanical activities go on are alive—are living; nor does he hesitate to call the activities vital.

"Now, it is not doubtful, as I see the matter, what it is, conceptually, that is working most strongly toward the laying of this old bogey. It is what has been for several decades forcing its way into technical biology under the name of the organismal or organismal conception. A highly significant thing that is happening at present in connection with this conception is its almost entirely independent growth in some half dozen subdivisions of the biological realm.

"The reach of the conception with its implications, into the sciences of not-living nature, into psychology, and into the domain of thought which has for many centuries been supposed to belong exclusively to philosophy, is certainly immense, though no one yet can tell exactly how far, or exactly what its effects will be.

"But what some of its most far-reaching consequences are to be is rather definitely foreshadowed. It will bring about a much sharper distinction between materialism and naturalism than has hitherto been grasped.

"Another result closely connected with this will be the recognition that man belongs wholly, i.e., with no aspect of his being excepted to the realm of nature.

"When the most distinctive characteristics of present-day civilization are viewed in the light of the psycho-biological matters touched upon here, I believe it must be concluded that man can never benefit to the full from the combination of his reason with his aesthetic, his ethical, and his religious emotions, until he recognizes these to be part and parcel of himself as a natural being."

And here is another statement:

"The most thoughtful biologists are Vitalistic or lean strongly toward it. As a matter of fact, some who call themselves Mechanists disprove it by their own writings. *A substitution of words does not convert the Vitalistic phenomenon into a mechanical phenomenon.* Several of our outstanding biologists who call themselves Mechanists, are really Vitalists by virtue of the very arguments offered in their writings, which in many instances are mere adaptations of the old scholastic arguments.

"One of the amusing instances of pretended Mechanism is that brought forth in the various articles on Emergent Evolution. This 'selective behaviour' of 'mass behaviour' referred to in these writings (though the authors themselves seem to miss this) has presented a new and one of the most powerful arguments for Vitalism as yet discovered; e.g., any drop of water will have the same behaviour all the world over, when given identical conditions. One *Paramecium*, as well as many another organism, does not behave precisely alike to all others, even under identical conditions. Many drops of water will still behave alike under the same identical conditions (freezing, boiling); but many *Paramecia* in a group or mass will behave startlingly different from the single individual. This is also shown in other species, e.g., the hunting of wolves singly and in packs, human beings under control of 'mob psychology,' etc."

The following letter throws some light on why many biologists feel that they must adhere to the laboratory findings exclusively and not permit themselves to be drawn into philosophical discussions even if these latter may really be of much greater importance to mankind than the mere adding of another isolated fact to an already large store of knowledge. Biologists, as well as all other laboratory workers, know that we can understand each other fairly well when we confine our speech to definite scientific terms but that we usually differ greatly when we enter into philosophical speculation. Scientists, therefore, have limited themselves to certain physically demonstrable methods.

Thus one biologist writes:

"My own observations of other persons' attitudes tempt me to say that as a rule they refuse to take Vitalism seriously, and I do not know whether this means that they have not thought about the possible limitations of science or whether, as in so many cases otherwise, they are ultra-practical as an accidental habit of mind.

"I am afraid many of us are more interesting as experimental animals for psychologists to study than as monuments of truth.

"My own view of this question is: 'Science limits itself voluntarily to what is reasonably demonstrable, using a perceptive method with logical extensions. As scientific individuals we have as much right to free and easy imagination, superstition, and the like as anybody else. We avoid these things for the sake of safety and security to our field of knowledge. I would think it almost equally unwarranted to state that either life is mechanism or that life contains a vital element which we do not or possibly cannot know. But science itself is a natural development of the human mind. Physics and chemistry are practical experiences, and whatever may be present in life, I am satisfied that so far these sciences have furnished the only kinds of facts out of which we could get any satisfaction, on the set basis of being perceptively observable, demonstrable, reasonable, commanding of universal respect, or of whatever quality scientific men desire to have as a criticism of truth.'"

REFERENCES

1. ALDRICH, J. M. The limitation of taxonomy. *Science*, 1927, **65**, 381-385.
2. BEACH, F. E. On the study of physics. *Pop. Sci.*, 1904, **65**, 61.
3. COCKERELL, T. D. A. Duty of biology. *Science*, 1926, **63**, 367-371.

4. DEWEY, J. In a discussion at the International Congress of Philosophy at Harvard University, Sept. 13-16, 1926.
5. KELLOGG, V. L. The biologist speaks of life. *Atl. Mo.*, 1921, **127**, 583-593.
6. MCCLUNG, C. E. The unity of life. *Science*, 1926, **64**, 561-569.
7. PEABODY, F. W. The care of the patient. *J. Amer. Med. Asso.*, 1927, **88**, 877.
8. PEARSON, E. L. Books in black and red. New York: Macmillan, 1923.
9. RITTER, W. E. Darwin's probable place in future biology. *Pop. Sci.* 1910, **76**, 32-48.
10. TORREY, H. B. Modern scientific thought and its influence on philosophy. *Pop. Sci.*, 1913, **82**, 85-99.

*Marquette University
Milwaukee, Wisconsin*

DEFINITIONS IN CHARACTER MEASUREMENT

JOHN N. WASHBURN

The predominant approach to the measurement of character recently has been by way of giving a numerical expression to a great variety of "traits"¹ and attitudes. Reverence, aggressiveness, honesty, caution, decision-speed, perseverance, self-assurance, studiousness, expansion-reclusion, introversion-extroversion, suggestibility, self-control, social perception, fair-mindedness, cooperativeness, emotionality, self-confidence, conformity, originality, impulse-judgment, and many other traits have been measured.²

For the most part each type of behavior has been evaluated independently. It is assumed by those who rate an individual in reverence that the presence of a considerable quantity of this trait is desirable, and a similar assumption is made by those who measure aggressiveness, conformity, originality, etc. There has been no common criterion of desirability.

Many of these traits are very subjective. They do not mean the same thing to different experimenters. Conklin (2), for instance, sees the "introvert" as described by Jung as one who evaluates the world in terms of its relation to himself. Allport (1, p. 117) also writes of the introvert as more egocentric than the extrovert, while others think of him as an individual who emotionally elaborates his experiences more than the extrovert, but not as one who is necessarily egocentric.³

¹The word "trait" is here, and throughout this article, used as being roughly synonymous with "characteristic behavior" or "function."

²Tests for most of these traits have been reported by Hartshorne and May (3). Subsequent work is reported in (4) and (5). See also (9) and (10).

³For example, Marston (7) speaks of "introversion being the dissipation of energy within the organism or its drainage through non-adaptive skeletal channels."

Aggressiveness may mean undesirable boldness to one investigator, or desirable initiative to another. Moreover, the same behavior may be very differently labeled, according to the investigator's notion of desirability, or according to the trait he happens to be seeking. A nursery school director rated a certain child high in "reverence" because she found him gazing fixedly for a long time at a flower. Another person might have called this behavior introversion, another persistence, or good attention-span, or dullness or interest, or curiosity, or love of beauty. In such an instance the trait name is determined not only by the subjective attitude of the observer toward the trait in question (i.e., by whether or not he is looking for it, or whether or not he considers it desirable), but also by his attitude toward the object which is stimulating the child. Thus, had the same child gazed with the same expression upon a snake he would almost surely have failed to make a score in reverence even though the investigator were looking for that trait.

It is clear that some traits can far more easily be observed objectively than can others. Where many people are reporting independently on a child they are more apt to agree on, say, temper tantrums than on reverence; there will be more agreement as to the validity of a test purporting to measure suggestibility than there will be as to the validity of a test of aggressiveness, or will-energy.

A chief reason for this is a great difference in clarity of definition. To illustrate with two of the examples we have used above, Webster says that reverence is "a mingling of respect, fear, and affection!" Funk and Wagnalls leaves out the element of fear, and the New English Dictionary leaves out both fear and affection. On the other hand, there is no disagreement as to the fact that suggestibility means "easily influenced by suggestion."

Much confusion could be avoided if people who attempt to measure a trait would state clearly what they mean by the trait name. A paragraph or more devoted to definitions in any work dealing with traits should be customary.

The same criticism applies to the measurement of character as a whole. Some think of character as practically synonymous with temperament, others think of it as practically synonymous with moral habits, still others identify it with will-energy, self-control, moral judgment, etc. To call such different things by the same name, leads to great confusion.

Nevertheless, some pioneers in character measurement consider definition unnecessary. "On what grounds," one of them was asked, "do you call these functions you have been measuring 'character' rather than, say, intelligence? What is your definition of character?" His reply was that this was an academic question: "We don't know what character is, but we are going ahead to measure it anyway—just as others have measured intelligence without agreeing upon a definition."

Let us follow the comparison further: It appears that in the case of intelligence, although there has been little agreement as to formal definition, there has long been general agreement as to the criterion by which marked proficiency or deficiency in the function can be judged. If an individual can get what he wants by indirect ("clever") methods he is considered intelligent. In spite of the fact that no one has said, "*Intelligence is the ability to reach an objective by indirection*,"⁴ nearly everyone, since Binet's time, has acted as if this were the definition. Hence it has been possible to dispense with a verbal formulation.

In the case of character there is agreement neither as to definition nor as to the validating criterion. There are many who, with the humanists, take as their criterion the suppression of desires and impulses in favor of an established moral code or principle. On the other hand, there are many who agree with J. S. Mill (8, p. 108) that "A person whose desires and impulses are his own—are the expression of his own nature, as it has been developed and modified by his own culture is said to have a character. One whose desires and impulses are not his own, has no character, no more than a steam engine has."

Many people apparently accept both of the foregoing standards, judging first by one and then by the other. There are, moreover, other standards that are occasionally encountered. Perhaps the chief of these is the standard of success. Sometimes people who have a combination of conative qualities which makes for achievement are said to have character.

Such a variety of standards makes definition either of the function to be measured, or the criterion by which it is to be gauged, imperative. One that is sufficiently ultimate to include an essential part of each of the principal existing standards will probably prove most serviceable. We may rephrase the three prevailing criteria as, roughly, (a) morality, (b) self-expression and integration (Mill's concept of integrated will and desire), (c) strength (i.e., conative factors contributing to achievement).

If we conceive of morality as that conduct which is believed to be essential to the greatest happiness (welfare) of others, and of integrated self-expression as that conduct which is essential to the greatest happiness (welfare)⁵ of the self, and of strength as the ability to control one's conduct in the interest of one's own happiness (self-integration) and the happiness of others (morality), an inclusive definition is at once self-evident: The degree of character is the degree to which one identifies one's own happiness with the happiness of others, and to which one can control and subordinate impulses which are contradictory to one's chief purpose.

⁴Köhler (6) points clearly to such a definition in his *Mentality of Apes* in the description of an "intelligence test" for his apes involving getting a banana out of a three-sided box by indirection.

⁵"Welfare" is put in parentheses as being subordinate to happiness, the latter being the more or less definitely ascertainable criterion of the former.

It may be pointed out that this definition can be paraphrased to read, "He has a superior character who loves his neighbor as himself, and who pursues what he believes to be the greatest good with all his heart and with all his soul and with all his mind."

There is nothing new in this concept certainly. But its implications in the field of character measurement have been overlooked.

This definition in its less Biblical form would, perhaps, be more readily accepted by some people if the term "welfare" were substituted for "happiness." But such an alteration would render the definition less serviceable, it would involve a criterion which is less measurable and which has more meanings than has "happiness." It is difficult, for instance, to conceive of even attempting to measure "welfare" as Goodwin Watson (11) has attempted to measure happiness among adult students of education. Nor is there apt to be as general agreement on what constitutes extreme cases of "welfare," whatever that may mean, as there is to be on what constitutes extreme cases of "happiness." Indeed, welfare has been a prime source of disagreement throughout the ages.

Let us say, then, that *every trait shall be evaluated by the contribution it makes to the economy of happiness*. Such a criterion is already popularly and loosely held. One who robs the needy is considered worse than one who robs the rich. One who cheats, say, the U. S. Postal Department is not so condemned as one who cheats to the same extent an individual. This is because the deeds appear to have different results so far as human happiness is concerned.

The popular view, however, is not consistent. Self-sacrifice is considered laudable even though the individual making the sacrifice loses more happiness than others gain. Such a view is the result of naïve logic: "Since extreme self-seeking is bad, and is among the most common causes of unhappiness, extreme self-sacrifice must be good and will result in general happiness." Or, "Since loving your neighbor as yourself results in happiness, loving your neighbor more than yourself will result in still greater happiness."

If, however, the criterion of economy of happiness is applied more reasonably, it is plain that there is no general increase in happiness when one person's loss is as great as, or greater than, another's gain. Such an application of the criterion will change many values. Spontaneous service, for instance, which involves ("selfish") pleasure in the happiness of others will take a higher place than dutiful and miserable self-sacrifice. In fact dutifulness, when its relation to general happiness is determined, may fall greatly in value, while the value of sympathy may rise.

Such values cannot, of course, be determined in advance. But the effort to do so in the light of any single, definite criterion will bring to light the need for certain tests which otherwise might be neglected. For example,

if the happiness of the group *and* the individual is accepted as the criterion, it at once becomes plain that it is necessary to discover the amount of happiness or unhappiness an individual may experience in simply being aware of these experiences in others, in short the need of a test of sympathy is revealed. This and allied traits, such as affectionateness, have been to date almost entirely neglected.

Finally, the economy of happiness serves a purpose that might, in fact, be served by any criterion that could be generally agreed upon and objectively determined. It makes possible the summation of various character traits. Sympathy, emotional stability, objectivity, and the many other traits of character may be conceived of as tendencies going in all directions. In order to evaluate these in comparable, quantitative terms, which can yield a meaningful sum, it is necessary to select one direction or trend and measure the others according to their divergence from or convergence upon the one selected. This will yield a number of plus and minus quantities, and the algebraic sum of the trends thus expressed is the character score.

Although any objective criterion makes possible the summation of various character trends in a single character score, it is by no means a matter of indifference what criterion is selected. If, for example, an individual's total character score should be in terms of "compliance," its importance for others than those in authority over him might well be questioned. Whereas, if the individual's score means the degree to which his qualities tend to promote his own happiness and the happiness of others, it has importance for all who come in contact with him.

SUMMARY

1. There is such disagreement as to the meanings of various trait names that any person attempting to measure a trait of character should state clearly his understanding of the trait name.

2. Traits of character are frequently not distinguished from traits of intelligence.

3. Although there is no generally accepted definition of intelligence, there is general, tacit agreement as to the criterion by which intellectual traits are evaluated—*the attainment of an objective by indirection*.

4. In the case of character there is neither a generally accepted definition, nor general tacit agreement as to the criterion by which character, or any character trait, can be evaluated. Among existing criteria are integration of personality, morality, and strength.

5. The three chief existing criteria can be harmonized if looked at from the point of view of the economy of happiness. Thus "integration of personality" becomes greater total personal happiness, "morality" becomes greater total group happiness, and "strength" the ability to pursue a greater happiness at the expense of a lesser. This leaves us with a single criterion and a definition. The criterion: *Character traits may be evaluated by the*

contribution they make to the economy of happiness, i.e., the total amount of happiness—one's own as well as the happiness of others. The definition: The degree to which one possesses character is the degree to which one identifies one's own happiness with the happiness of others and to which one can control and subordinate impulses which are contrary to one's chief purpose.

6. Such a formulation, aside from causing a re-evaluation of certain traits which are now apparently valued for themselves alone, makes the determination of character far more objective than would be the case if terms such as "welfare" were employed, and calls for extensive statistical study of happiness and the factors contributing to it.

7. Any concise formulation of a criterion makes possible comparable quantitative expressions for individual character traits in terms of their departure from or approach to the criterion. These individual departures and approaches may be summated in a single significant character score. The more generally the importance of the criterion is felt, the more significant will be the character score.

REFERENCES

1. ALLPORT, F. H. Social psychology. Boston: Houghton Mifflin, 1924. Pp. xiv+453.
2. CONKLIN, E. S. The definition of introversion, extraversion and allied concepts. *J. Abn. & Soc. Psychol.*, 1922-1923, **17**, 367-382.
3. HARTSHORNE, H., & MAY, M. A. Personality and character tests. *Psychol. Bull.*, 1926, **23**, 395-411.
4. ———. Studies in deceit. Book 1. General methods and results. Book 2. Statistical methods and results. New York: Macmillan, 1928. Pp. xxi+414; ix+306.
5. HARTSHORNE, H., MAY, M. A., & MALLER, J. B. Studies in service and self-control. New York: Macmillan, 1929. Pp. xxiv+560.
6. KÖHLER, W. Mentality of apes. (Trans. by E. Winter.) London: Kegan, Paul, 1924. New York: Harcourt, Brace, 1925. Pp. viii+342.
7. MARSTON, L. R. The emotions of young children. *Univ. Ia. Stud.: Stud. Child Welfare*, 1925, **3**, No. 3. Pp. 99.
8. MILL, J. S. Liberty. London, 1859.
9. MOORE, H. T., & GILLILAND, A. R. The measurement of aggressiveness. *J. Appl. Psychol.*, 1921, **5**, 97-118.
10. WASHBURN, J. N. An experiment in character measurement. *J. Juv. Res.*, 1929, **13**, 1-18.
11. WATSON, G. B. Happiness among adult students of education. *J. Educ. Psychol.*, 1930, **21**, 79-109.

Syracuse University
Syracuse, New York

SEX DIFFERENCES IN CONVERSATIONAL INTERESTS

STUART M. STOKE AND ELMER D. WEST

In 1922 Moore (3) lamented the fact that intelligence tests and other standardized forms of studying individual differences had shown no sex differences of significance. He was convinced that such differences existed and supported his views with data collected from listening to fragments of conversation on Broadway during the evenings. One hundred seventy-four such fragments were overheard. In 80 cases the conversations took place between men, in 30 between women, in 32 a man was speaking to a woman, and in 32 a woman was speaking to a man. From this study he concluded that there were: "... very considerable and ineradicable differences in the original capacities of the two sexes for certain types of enthusiasms. . . . After making every possible allowance for differences in convention and personal experience, it is hard to escape the conviction that the original nature here depicted is of two fundamentally different sorts" (3, p. 214).

Landis and Burt (2) were impressed with the results but felt that a sufficient allowance might not have been made for the type of place in which the conversations took place, and also that the number of fragments overheard was insufficient. Therefore they undertook to repeat the study in a different locality, Columbus, Ohio, to collect conversations in many places, and to account for social status and occupation as far as possible. They listened to fragments of conversations in a great variety of places, obtained a considerable number of such fragments, and concluded that "the results are in marked agreement with those of Moore" (2, p. 83).

Landis (1) repeated Moore's study in a certain locality of London and reached much the same conclusion as his predecessors. Under the conditions described there seems to be a marked sex difference in conversational interests. The main differences are disclosed in Table 1, adapted from Moore (3).

The sex differences shown in Table 1 are marked indeed, but the writers are disposed to believe that insufficient attention has been given to the variables of occupation, location at the time of the conversation, the type of person with whom the conversation was held, age, and social status. The following study attempts to remedy these defects as much as possible by

TABLE 1

Topics discussed	Men to men percentages	Women to women percentages
Money and business	48	3
Amusement	14	4
Persons of the same sex	13	16
Clothes, buildings, interior decoration	2	23
Persons of the opposite sex	8	44

holding constant the following factors: (*a*) occupation—only college students were used; (*b*) age—probably age differences seldom exceeded a four-year range; (*c*) time of conversation—at night; (*d*) type of persons among whom the conversation was held—only one sex and only college students present; (*e*) place—the residences of the students where the greatest amount of freedom would be expected; (*f*) social conditions—records for sorority and dormitory women were kept separate. Only a few non-fraternity conversations occur among the records of the men. Since these did not differ materially from the records of the fraternity men, they were all treated together.

The conversations took place between March, 1929, and February, 1930. Thirty-six students participated in gathering the data. Each was furnished with a check list of probable topics and asked to check those mentioned during the "bull sessions" which they attended. Topics not listed were to be recorded and in each case the date of the conversation and the number of the participants noted. Each recorder was instructed to say nothing about his work but to record the topics after he had left the group. No doubt more topics were discussed than recorded, due to the inaccuracies of memory, but it is probable that these omissions were scattered somewhat evenly over the list. It is unlikely that any selective factor of importance operated through memory to obscure results. All types of recorders were used, ranging from Phi Beta Kappa students, on one extreme, to those who were failing courses, on the other. Only one kind of recorder was studiously avoided and that was the sort which dominates conversations. The records of such a one would be those of his own interests only. Time spent upon the various topics could not be secured without betraying the fact that a study was in progress, and therefore it was not attempted.

The study covered practically a college year. During this time a total of 498 sessions devoted to "talk" were recorded. Of these, 239 were held by the women and 259 by the men. A total of 2230 topics were recorded—1353 being reported by women and 877 by men. The 1353 reported by women were divided between the sorority and dormitory groups—747 being reported by the former and 606 by the latter. The average number of topics discussed per session by the men was 3.4, while the average number by women was 5.7. (There was little difference between the two groups of women.) This is a rather marked difference and three possible causes suggest themselves. First, that women hold longer sessions of conversation—a doubtful assumption. Second, that the women were better recorders than the men. This seems more likely, but that all of the difference is due to superior recording on the part of the women is doubtful. The third explanation is that men dwell longer upon a topic than women do. The writers are inclined to accept both the second and third reasons as partial explanations of the difference noted.

The range of topics differed little. The topics which appear exclusively

TABLE 2

TOPICS DISCUSSED BY COLLEGE STUDENTS AND THE PROPORTIONAL
DISCUSSION OF EACH TOPIC IN DIFFERENT STUDENT GROUPS

Topics	Percentages		
	Fraternity members	Sorority members	Dormitory women
Dates	6.8	6.3	9.9
Fraternities	5.7	1.7	3.3
Sports	4.8	2.4	1.6
Dancing	4.6	4.4	3.6
Clothes	4.2	3.6	6.0
Drinking	4.0	2.2	1.1
Campus politics	3.8	2.5	1.5
Money	3.4	2.8	4.4
Week-end trips	3.4	3.5	2.0
Studies	3.1	1.6	3.0
Summer employment	3.1	1.7	2.0
Food	2.9	4.5	3.9
Sororities	2.9	5.9	3.3
Grades	2.9	2.0	.7
Stories (shady)	2.4	2.1	2.0
Gossip	2.3	4.0	4.9
Sex problems	2.2	2.5	4.9
Vocations	2.2	1.6	1.3
Religion	2.1	1.4	2.8
Vacations	2.1	2.8	1.3
Music	1.9	2.0	1.9
Airplanes	1.8	.1	.1
Home and relatives	1.8	2.2	4.4
Smoking	1.8	1.9	1.5
Examinations	1.8	1.2	.5
Teachers	1.7	4.1	2.8
Kissing and necking	1.7	2.5	2.9
Shows	1.6	2.8	1.0
Comparisons of universities	1.6	.6	1.6
Inventions	1.3	.4	
Automobiles	1.3	.6	2.6
University regulations	1.3	1.2	2.6
Home towns	1.1	1.9	1.3
Debates	1.0	.5	.1
Politics (not campus)	1.0	1.3	
Humorous stories	1.0	1.2	.8
Stories (neither humorous nor shady)	1.0	1.9	.5
Travel	.9	1.3	.3
Drama	.8	1.4	1.1
Sedentary games	.8	.9	.3
Literature	.7	1.4	2.0
Culture	.6	2.0	.7
Hazing	.5	.2	.3
Gov't ownership of utilities	.3		
Animals	.2		
Art	.2	1.2	.1

TABLE 2 (*continued*)

Topics	Percentages		
	Fraternity members	Sorority members	Dormitory women
Death	.2	.5	.7
Serenades	.2	.6	.7
Dreams	.2		
Social conventions	.1	1.1	1.6
Nature appreciation	.1	.8	.5
Evolution	.1	.4	.1
Socialism	.1	.1	
War	.1		
World peace	.1		
Chain stores	.1		
Women	.1		
Marriage		.1	.3
Personalities		.1	
Emotions		.1	
Fraternity pins		.1	
Letters		.1	
Moving picture actors and actresses			.1
University life			.1
Joy and problem of being young			.1
Accidents			.1
Speech provincialisms			.1
The ideal man			.1
What men like for a girl to be			.1
Ideals and ambitions			.1
Love			.1

in the lists by men or by women are discussed only from one to four times, and hence no topic can be regarded as the exclusive property of either sex.

Table 2 shows the percentage of the total conversations of each group which each item furnished. For example, dates comprised 6.8% of the conversations of the men, 6.3% of the sorority members' conversations, and 9.9% of the dormitory women's. Scanning the columns even casually shows that the differences between sorority and dormitory conversations are frequently greater than the differences between sorority and fraternity groups. In order to reduce these differences to a single unit of comparison, the writers calculated the sums of the differences regardless of sign between (a) the fraternity and sorority percentages, and (b) the sorority and dormitory women's percentages. The sum of the differences between fraternity and sorority groups for 47 mutual topics is 35.8, while the sum of the differences between the dormitory women and sorority members for the same topics is 45.5. Thus it is evident that a social factor of selection has more effect upon conversational interests than sex does, when the results are treated in this rough way.

If certain topics with a central theme are selected from the general list,

TABLE 3

THE FREQUENCY OF DISCUSSION OF TOPICS OF SEX INTERESTS IN
DIFFERENT GROUPS
(Percentage basis)

Topics	Fraternity members	Sorority members	Dormitory women
Dates	6.8	6.3	9.9
Dancing	4.6	4.4	3.6
Clothes	4.2	3.6	6.0
Kissing and necking	1.7	2.5	2.9
Sex problems	2.2	2.5	4.9
Stories (shady)	2.4	2.1	2.0
Women	.1		
Marriage		.1	.3
The ideal man			.1
What men like for a girl to be			.1
Love			.1
Totals	22.0	21.5	29.9

TABLE 4

FREQUENCY OF DISCUSSION IN DIFFERENT GROUPS OF ITEMS DEALING
WITH PERSONALITIES
(Percentage basis)

Topics	Fraternity members	Sorority members	Dormitory women
Gossip	2.3	4.0	4.9
Home and relatives	1.8	2.2	4.4
Teachers	1.7	4.1	2.8
Personalities		.1	
Joy and problem of being young			.1
Movie actors and actresses			.1
Totals	5.8	10.4	12.3

we find further evidence of the strong effect of environment in determining conversational interests. In Table 3 there is a group of topics centering around the subject of sex. The fraternity groups discussed these things 22% of the total number of topics they mentioned, the sorority groups talked about them 21.5%, but the total of the women's dormitory groups is 29.9%. The writers suspected that the excess of the dormitory women over the sorority women was due to insufficient social life, but in order to check their opinion they asked two classes in psychology which group had more dates, more clothes, more social life, etc. The classes were not apprised of the reasons for asking the questions. The votes were about two to one

that sorority women have more of these things. Obviously we have no sex difference in interest shown here, but just another example of conversation being determined by thwarted desires. In previous studies women were reported as discussing the topic of clothes quite extensively while men seldom mentioned it. Here we find that the young male is equally interested in the art of self-decoration when we hold the factor of social background constant.

Table 4 shows the frequency of discussion of topics dealing with personalities. Such items as are closely related to the sex interests just discussed were omitted from this table in order that the environmental factor just noted might not influence this finding also. It may be possible that the dormitory women discuss home and relatives so much more because of less social life at college and consequently greater tendency to homesickness. However there is evidence in Table 4 of a sex difference, although it is not as great as the differences previously reported. For example, Moore's study shows women talking to women about people in 60% of the topics overheard, while men talked to men about people in only 21% of the fragments overheard. In this study we find personalities forming only 5.8% of the things discussed by men, 10.4% of the things discussed by sorority members, and 12.3% of the topics mentioned by the dormitory women. Sex differences in interest in personalities do exist according to all of the studies of conversational interests, but it is apparent that, when the many variables are held constant and the conversers are young, the sex difference is reduced markedly. Whether the amount of sex difference in college students, with respect to interest in personalities, is native or acquired cannot, of course, be determined from this study. However, there has been plenty of time and probably influence, also, for producing such differences long before college age is reached.

Women are somewhat more interested in cultural topics than are men. If "music, travel, drama, culture, art, nature appreciation, and literature" are grouped together we find that 5.2% of the men's topics are included, 10.1% of the sorority topics, and 6.6% of the dormitory women's. Since the fraternity and sorority groups have much the same social background, it is evident that when this factor is held constant a sex difference appears. That this difference is native is doubtful. At any rate, we discover almost as large an environmental difference between two groups of women as we find between the two sexes where social traditions are fairly constant.

Men are more interested in sports than women, although again we do not find as marked differences as have been previously reported. The sorority girls discuss sports more than the dormitory women, which is probably due to the influence of more contacts with men. Men discuss drinking a great deal more than women do—probably because it is largely a man's problem in this particular environment. Smoking is about equally

discussed by all groups. Approximately the same percentage of "shady stories" appear in each group, although it is quite possible that there may be differences in defining the term. At any rate, however, the women violate their own code of good taste in this respect about as frequently as the men. University regulations are discussed, in point of frequency, in the same order as they affect the groups. Men discuss them but little, the sorority women more, and the dormitory women most. Women do show greater interest in social conventions than men, but in view of the much greater amount of training which women receive along this line, one would scarcely be justified in hailing it as a true sex difference. Men do talk more about "things" such as inventions, airplanes, etc., than women, but the differences are not as large as expected.

CONCLUSIONS

Where the variables of sex, age, race, occupation, social background, and place of conversations are held constant, the sex differences in conversational interests are rather small. Men still show more interest in "things" and in sports, while women show more interest in personalities, cultural topics, and social conventions. Other occasional differences seem traceable to environmental influences. It is possible that men tend to pursue a given topic longer than women do, but positive proof of this is lacking. Neither men nor women of college age seem to be burdened with serious conversations.

REFERENCES

1. LANDIS, C. National differences in conversations. *J. Abn. & Soc. Psychol.*, 1927, **21**, 354-357.
2. LANDIS, M. H., & BURTT, H. E. A study of conversations. *J. Comp. Psychol.*, 1924, **4**, 81-89.
3. MOORE, H. T. Further data concerning sex differences. *J. Abn. & Soc. Psychol.*, 1922, **17**, 210-214.

Ohio University
Athens, Ohio

A NOTE ON THE ATTITUDE OF SOCIAL CONFORMITY

PAUL R. FARNSWORTH AND ALICE BEHNER

In his *Social Psychology* (Houghton Mifflin, 1924, p. 278) Floyd Allport has stated: "Barring individual exceptions . . . there is a basic human tendency to temper one's opinions and conduct by deference to the opinions and conduct of others. . . We may call this the *attitude of social conformity*." Allport has based this conclusion in part on an experiment in which 15 subjects estimated the weights of 10 objects, identical in appearance, in relation to a light and heavy standard. The subjects carried

on their estimations twice—once individually, and once in the presence of their fellows. "When judging in the group the heavier weights were judged as *lighter* than when judging alone; and the lighter weights were judged as *heavier*."

Although the writers are not certain that these experimental results indicate an attitude of social conformity, they seemed of sufficient importance to warrant experimental verification. As Allport's data were all given in graphic form, it appeared desirable to ascertain mean differences, sigmas, and the reliabilities of the mean differences. Somewhat larger populations (43 students in a psychology class) could be and were employed in the present study.

The directions given were as follows: "You are to estimate the weights of ten bottles relative to the weights of two samples. Consider the top horizontal line to represent the weight of the heavy sample, and the base line the weight of the light sample. For example, if a certain bottle seems midway in weight between the light and heavy samples, check its line at the midpoint. None of the bottles will weigh as much as the heavy sample nor as little as the light sample. Bottle 10 weighs more than 9, 9 more than 8, etc. Each bottle has a line on which the weight is to be estimated."

The rating lines were 127 mm. in length. The heavy samples weighed 120 g. and the light 12 g. The weights of the stimulus bottles (7 cm. long, 2 cm. in diameter, and filled with cotton and shot) are given in Table 1. In the "alone" portions of the experiment the subjects tested themselves with no one else in the room. The "in class" portion was held in a classroom in which were 43 students each of whom was supplied with a set of 12 bottles. The testing was in the order: (a) a group of 24 subjects tested

TABLE 1

Actual No.			A		B		C		D	
	<i>g.</i>	<i>mm.</i>	24 "alone"	19 "in class"	43 "alone"	43 "in class"	<i>M</i>	σ	<i>M</i>	σ
			<i>M</i>	σ	<i>M</i>	σ	<i>M</i>	σ	<i>M</i>	σ
1.	15.0	3.5	5.3	3.7	7.4	3.1	5.3	3.3	6.2	3.2
2.	18.8	8.0	12.2	6.5	14.6	5.5	12.1	5.9	13.2	6.2
3.	23.5	13.5	20.2	8.3	23.8	8.6	20.1	7.8	21.9	9.0
4.	29.3	20.3	30.0	11.8	34.2	12.2	29.8	10.6	32.0	11.7
5.	36.6	28.8	40.1	12.3	46.1	14.3	41.5	11.9	43.5	13.6
6.	45.8	39.6	50.7	12.2	57.8	15.4	53.5	12.8	55.6	14.1
7.	57.2	52.9	63.7	12.8	71.8	16.6	67.0	13.9	69.4	14.3
8.	71.5	69.6	79.5	14.4	88.0	14.3	82.6	14.5	85.7	13.8
9.	89.4	90.6	97.4	12.7	101.2	13.1	99.8	11.8	100.6	12.1
10.	111.7	116.7	118.7	5.1	117.1	8.0	118.8	5.1	117.1	8.1

g refers to the weights of the stimulus bottles in grams, *mm.*, to their rating equivalent in millimeters, *M* to the means of the estimates, and σ to the sigmas of the distributions (in millimeters).

individually over a period of 3 weeks; (b) the 24 plus 19 others tested together, making a retest for the 24 but an initial test for the 19; and (c) the 19 in a retest over another period of 3 weeks. Comparisons have been made between the groups of 24 and 19 in their initial tests (the former "alone" and the latter "in class") and between the "alone" and "in class" portions of the entire 43.

It is apparent from the above data that the mean scores are essentially similar to those of Allport's experiment. In Column A ("alone") the first mean is smaller (more extreme) than is the first mean of Column B ("in class"). $\frac{\text{Diff.}}{\sigma \text{ diff.}}$ is 2.1 (98 chances in 100)—a difference which many sta-

tisticians consider to be statistically reliable. However, at the other end of the series (last mean) there is a $\frac{\text{Diff.}}{\sigma \text{ diff.}}$ of .8 (79 chances in 100) which is

in the correct direction to verify Allport, but not of sufficient magnitude to be very significant. In comparing Columns C and D, the $\frac{\text{Diff.}}{\sigma \text{ diff.}}$ formula

was modified in the usual manner to take the factor of correlated data into account. The first mean of the "alone" group is again seen to be smaller than the first mean of the "in class" group. The value here was 1.7 (96 chances in 100). At the opposite end of the series the value was 1.3 (96 chances in 100). Both of these figures are lower than might be desired to perfectly substantiate Allport's findings, but are in the proper direction. Forty-two per cent of the subjects had curves similar, at the ends, to the composite curves. That is, the means of the "alones" were smaller at the lighter end of the series and larger at the heavier end. Allport has reported that 66 per cent of his subjects followed the composite curves. It is to be noted that the subjects overestimated the weights in almost every instance. This was done more "in class" than "alone" (an exception existed in the case of the heaviest weight). Another interesting item in the data was the fact that the sigmas did not get progressively larger with the heavier weights.

This experiment has verified, to a certain extent, Allport's previous findings that, "When judging in the group the heavier weights were judged as *lighter* than when judging alone; and the lighter weights were judged as *heavier*." At least the *heaviest* weight was judged as *lighter*. Somewhat similar mean scores, then, have been found. However, the statistical significance is not so clear-cut as might have been desired.

Stanford University
California

FIRST INTERNATIONAL CONGRESS OF RELIGIOUS PSYCHOLOGY

The First International Congress of Religious Psychology will be held at the University of Vienna from May 26th to May 31st (Week of Whitsuntide), 1931, under the auspices of the International Society for Religious Psychology, the President of which is Karl Beth, D.D., Ph.D., Professor of Protestant Theology in the University of Vienna, and the Vice-Presidents, Franz Brandl, LL.D., Director of the Viennese Police, and Otto Nahrhaft, LL.D., First Public Prosecutor.

The most important problems of recent religious psychology will be discussed from various points of view, educational, experimental, sociological, psychiatric, pathological, theological, occultistic, etc., the main subject being the psychic basis of the religious unbelief of the present day.

Many internationally known scientists are scheduled to deliver lectures or read papers at this Congress.

Opportunities to see the city of Vienna and its important institutions will be amply provided for.

Inquiries as to program, facilities, schedule of sightseeing trips, and all other information should be sent to the International Society for Religious Psychology, Vienna VII. Zitterhofergasse 8, Austria, and should be accompanied by fifty cents in stamps.

BOOKS

MARGARET MEAD. *Coming of Age in Samoa*. New York: Morrow, 1928. Pp. xv+297. \$3.00.

MARGARET MEAD. *Growing Up in New Guinea*. New York: Morrow, 1930. Pp. xii+372. \$3.50.

It was early seen that ethnology and psychology are basically the same discipline; but nothing was ever done about it. Miss Mead's contribution consists in having not only seen it, but reacted to it—fruitfully; in this she is accompanied, I think, only by Malinowski. The core of her reaction is the methodological one of using environment (in a fundamental sense, not the superficial one concerned with father's occupation and number of books in home library) as the principal variable. At one stroke this simple idea goes far toward throwing indispensable light on two long-standing obstacles—one the uncontrolled variation of the "unconscious," occasionally pointed out by psychoanalysts in criticism of experiments in academic intellectualist psychology, the other the fact that the greatest environmental variations ordinarily procurable in heredity-environment studies are so slight as to make correlated differences in performance uncomfortably small also. It need hardly be said that the advantage is accompanied by disadvantages—the conditions of observation are heterogeneous, precise measurements are practically impossible, the investigator's personal bias cannot be eliminated or allowed for; heredity is largely an unknown quantity — though in the Admiralties the social system permitted some use of the foster-child device.

But it may be well to recount briefly the technique and content of the observations. Miss Mead worked in each case with villages of sea peoples in the Oceanic tropics; in each she spent several months in daily contact with the people, conversing with them in their own language, and devoting herself chiefly to the psychology of the women and children. The intensive nature of the contact is worth stressing; one can count on the fingers of one hand the prominent anthropologists who have spoken to their subjects habitually in the latter's own tongue; but when the list is further limited to those who have lived among them on equal terms and made a definite effort to experience their problems and attitudes, it dwindles, I believe, to Malinowski and Miss Mead. Among the brown villagers of Samoa she found a highly communal society with slight drive and conspicuous lack of conflict, but considerable imagination and capacity for harmonious living. The blacks of the Great Admiralty were about as different as they could have been—a grasping, business society without embellishment, with a high degree of drive and the corresponding anxieties and quarrelsomeness. These differ-

ences were reflected, of course, in the education and satisfactions of the children, topics which formed the investigator's primary concern.

One cannot forbear to underscore Miss Mead's concept of the pageant of human societies as essentially a series of great sociological experiments, involving thousands of individuals and lasting hundreds of years—a series which the European is eradicating with characteristic speed and efficiency, so that it is imperative in a peculiarly poignant sense to read as much as we can of the records of these experiments before the imminent day when they can teach us no more. It is so that she has conceived her own work, and in the same terms we may attempt to formulate a somewhat crude scheme for presenting in an orderly fashion the many facets of her results. We may generalize the problem of the Samoa volume to read: Is personality as we know it inherited or acquired? This is a psychological problem, and the answer to which her results lead her is: Acquired, predominantly. On this basis the Melanesia volume may be generalized (though no longer as a simple dichotomy) somewhat as follows: Given three (acquired) personality results (one American and two Oceanic) and the processes of their acquisition, what modifications of the latter are indicated as likely to condition desirable modifications of the former? This is an educational problem; its answer is not simple, but may be formulated approximately thus: Opportunity for identification with forceful, integrated, and cherished personalities, as with the Manus; opportunity for assumption of responsibility proportioned to capacity, as in Samoa; and a tradition potentially rich, as in America.

Among the psychological questions subsidiary to the one stated above, that of the classical disturbances of adolescence evidently ranked first in Miss Mead's interest, for it was this problem which sent her to Samoa to seek the answer by the environment-variate technique. The answer was unequivocal, also as indicated above: The mental conflicts of adolescence are not consequences of its physical changes, but are the results of the interaction of these with the training and other cultural environment. A close second to this in interest is the related problem of family relationships, particularly with respect to the psychoanalytic findings. It will be recalled that Malinowski, who early regarded the formulations of Freud as highly significant for anthropology, later altered this position on the basis of his observation that among the matriarchal Trobrianders conflicts centered about the maternal uncle and the sister instead of the father and mother. Miss Mead found additional variations of the greatest theoretical interest; in Samoa the filial relationship—due to the prevalence of the clan type of household—is diffused over a number of adult relatives; furthermore, the relationship lacks "teeth," as an aggrieved child can at any time change his residence to a more congenial household; still further, the child himself occupies a position of authority toward an ever-growing number of juniors. Among the

Manus several additional factors complicate the picture: marital relationships are openly unpleasant; no demands whatever are made on the child except those necessary for physical safety—he is free at any time openly to defy any adult with impunity; and a peculiarly tender relationship toward the father exists throughout the formative years and is reflected in a marked temperamental resemblance of children (and foster-children) to their fathers. Miss Mead herself, in her *Anthropologist's Footnote to 'Totem and Tabu' (Psychoanalytic Review)*, indicates the amplified generalization to which these variations seem to point, viz., the impulses (hate, love, etc.) are innate, the classical complexes are manifestations of their conflict with cravings for approval (love) from the group—which cannot be granted given certain cultural patterns. Thus the Oedipus conflict is of slight intensity in Samoa, since there is no tradition of disloyalty or delinquency upon desertion; and its father-hatred component is absent in Peri because the father may be flouted at will while still granting his love; in the Trobriands the picture is a simple transposition of dramatis personae. It would be well if *Totem and Tabu*, particularly, could be rewritten in these more generalized terms; that its thesis occurs as a widespread fantasy in our civilization no one with analytic experience will question, but it is couched as anthropology, and, so interpreted, it represents the most seriously dissonant note in the authoritative writings on psychoanalysis; that the venerable author has apparently admitted as much in ambiguous terms, while continuing to retain the original form in his theoretical writings, does not help matters.

A third matter of interest in the nature-nurture department is that of the development of fantasy. The Samoan pantheon is limited approximately to the ghosts that haunt the palm groves—not too hostile a crew—and the malevolent spirit of the unborn child; there seems to be almost no development of myth and legend; to these types of fantasy should perhaps be added the rather extensive development of rank. To be correlated with this slight growth of compensatory activity is the paucity of demands of all kinds on the individual—the economic demand, the social demand, the ethical demand. Manus society is sharply stratified on an age basis, the demands on the children being much less than in Samoa, those on the adults much more; the fantasy development runs parallel. There are no stories, no building or other representation, among the children, and the spirits of the adults are but half credited; but in the adult world, ridden by anxieties, obligations, and the perpetual necessity for “climbing,” the spirits, grown strict, somewhat malevolent, and self-indulgent since death, play a directing part. The tremendous part of fantasy in our Western world that thwarts more than it satisfies is too familiar to require comment. Though the series contains but three instances (it might easily be amplified), it appears to be an entirely valid instance of the method of concomitant variation.

The light shed by the three cultures upon the developmental mechanisms

of personality is considerable. Identification appears to play the most prominent part, and so the entire genesis of personality links up with the "Oedipus" situation. The Samoan personality is attractive, but largely undifferentiated and superficial; there is also little interest in psychological phenomena—most of them are settled by being (very generically) named. Correlated with this is the almost complete lack of privacy from birth—the constant association with groups of ten to fifty people of all ages and both sexes; the identification is thus with a sort of general average. The Manus child is less a bit of public property; his principal companion from weaning to seven or eight is his father, after that for several years older children of the same sex. Accordingly the children (and foster-children) of aggressive established men are aggressive, of unaggressive men are unaggressive, with those of aggressive unestablished men in an intermediate position. It is not easy to say anything simple about the incredible complexity of American personality and its determinants; insofar as it shows any one thing, it is possibly the influence of a standardized inconsequential female school-teacher, with a remote flavor as of a male week-end visitor, somewhat irritable, who is understood to furnish the finances.

Sex attitudes are of course determined by social pressures; the determination of these pressures in America, while complicated, is with fair probability to be ascribed to "Oedipus" anxieties working historically through theological channels; in Samoa they seem, also fairly simply, to derive from the unimportant place assigned to personality, and so ultimately from the patterning of the individual after the average—necessarily of less content than any individual. But in Peri it is not easy to account for the ingrained puritanism, so like our own calvinisms. The one thing certain about it is that it is a taught repression, dependent for its maintenance on threatened withdrawal of love.

The group of educational findings and problems cannot be so readily set forth in order; but one feels that (contrary to the usual habits of investigators who draw practical proposals from their findings) much of the importance of Miss Mead's work lies here. This material is collected in Part II of the New Guinea volume, and is based on the findings with both groups; the expressive chapter title "Bequeathing our Tradition Graciously" conveys the burden of the argument, since the author points out that, so far as effectiveness is concerned, the only conclusion that can be drawn from the findings is that one method leads to about the same final result as another—even the Manus rowdy youth who jeers the mores of his elders being reduced by the force of circumstances in a few years to anxious and cowed compliance with their demands. But the effects upon personality and the sum total of happiness in the group bear quite other implications, as does also, no doubt, the net progress of the group in the universal human task of mastering the environment. The outstanding desiderata to which

the findings lead are the preservation of the gains of one generation by its successor, the necessity for developing a coherent culture before it can be transmitted, the desirability of training by responsible participation, and accent on richness of personality instead of appurtenances.

The first of these considerations leads directly, of course, to the reflections on identification mentioned above; it has as a necessary corollary the application of the same idea to the other end of the scale—in a system which permits the forceful personality of a leader to be transmitted to his child we must be prepared to have the inadequacies of the rank and file similarly transmitted; the only alternative is to make the forceful individual a teacher, entrusted with influencing not only his own child's personality, but those of his weaker neighbors as well. There is no concrete probability of anything of this sort coming to pass in our culture, in which lip-service to education is combined with underlying contempt; nor, for that matter, is there likelihood that we shall soon develop a culture in which any appreciable number of forceful males will have the leisure and the inclination to associate intimately with their children. The actual prospect, for America, seems to be the increasing homosexualization in greater or less degree of the great majority of our male population, a process the results of which are already perceptible in any examination of the mental hygiene of the community. For Peri, the system seems to be dynamically one in which differences become greater, with a constant elimination of the unfit as they fall below the competitive level (with occasional reversals, of course). For Samoa, it seems unlikely that any cultural or personality advance can be maintained, since no mechanisms for perpetuating it exist.

The perplexing question of sub-cultures can hardly exist in the primitive, simple cultures studied by our author; they are, however, a constant element in our own society and education. The author points out the disaster which may overtake, for example, the child habituated to a naturalistic sexual outlook; and indicates that it was necessary for the Russian experiment to be assimilated throughout a coherent cultural group before it could be transmitted by education. But the modern scene has become such a complex of sub-cultures that conflict can hardly be avoided, and their disintegration has proceeded everywhere with an inevitability ascribable only to the insistence of the immediate major drives. It seems, in fact, that this consideration, while theoretically correct, is practically barren; cultural coherence can rest on only a few bases—the elemental wolf-pack idea, the mobilization of anxieties and fantasy life, and renunciation-sublimation; on the basis of the second we have partially passed the first, but advance to the third seems a bootstrap maneuver now nowhere above the horizon.

Induction into a culture by responsible participation proportionate to capacity is, on the contrary, just around the corner, if not already arrived in the most advanced communities. It is called by the inadequate title of

"progressive education," and while in all too numerous cases the idea has been distorted to provide a reputable mask for ancient forms of adult domination, nevertheless there is a widespread and genuine movement in this direction. This stands, I believe, as one of the very few respects in which (from the point of view of human values) our own culture has developed a trait clearly superior to the corresponding one in both the primitive societies under consideration.

There is little hope, however, for the adoption or even approach to the substitution of character for possession values in our society, as with the Manus. It is difficult to see this as other than a consequence of crude economics; and if this ascription be correct, the possibility of such an alteration on any general scale practically vanishes. The superior achievement of the Samoans in this respect seems to be due to the fact that there is enough to go around—it is of no serious consequence whether the neighbor or the chief is honest (and no reason for him to be otherwise) since a few hours' labor secures the future. But the Manus live over the lagoon; fish alone are inadequate for existence, and in trade the only certain road to security is the possession of plenty of dog-tooth currency—and conversely, the possession of the same by another is sufficient cause for esteem, i.e., propitiation. So in America; he that is without power may be estimable, in the abstract; but in a world which demands the diversion of such a disproportionate share of one's energy to self-preservative reactions toward those who have power, the interest which remains free for attachment to him dwindles or disappears.

Enough has been said to give a brief idea of the content and significance of these stimulating volumes; in closing, a word from the author herself regarding the relationships of her activity as a whole can hardly be improved: "The ethnologist. . . says to the psychologist who has made a long and careful investigation within our society, from which he may or may not have drawn conclusions which he regards as final, 'Let me take your results and submit them to a new test. You have made such and such generalizations about the thought content of young children, the relationship between mental and physical development, the connection between a certain type of family life and the possibility of a happy marital adjustment, the factors which go to the formation of a personality, etc. These results I find significant and important. Let me therefore submit them to the test of a different social environment, and in the light of such observation, on the basis of our combined research, on the basis of your initial definition of the problems and observations within our society, and my check observations within a different society, come to conclusions which will successfully withstand the accusation that the effect of social environment has not been properly allowed for. It will then be possible for you to divide your observations upon individuals within our culture into two parts: data upon the

behavior of human beings modified by present-day culture, which will be of the utmost importance in handling educational and psychiatric problems of individuals within the same cultural background, and second: theories of the original nature, the potentialities of man, based upon your observations and mine.'... Meanwhile the psychologist should offer suggestions for research. Many field trips which are now only historical investigations, of extreme value in adding to our knowledge of human society and the lengths to which it can influence human behavior, are only half as valuable as they might have been if definite psychological problems could be attacked simultaneously."

RAYMOND R. WILLOUGHBY

Clark University
Worcester, Massachusetts

- BEARD, C. A., & BEARD, W. The American Leviathan: the republic in the machine age. New York: Macmillan, 1930. Pp. xv+82+. \$5.00.
- FLÜGEL, J. C. The psychology of clothes. (Int. Psycho-Analytical Library, ed. by E. Jones, No. 18.) London: Hogarth Press & Institute of Psycho-Analysis, 1930. Pp. 256.
- GARTH, T. R. Race psychology: a study of racial mental differences. New York: Whittlesey House (McGraw-Hill), 1931. Pp. xiv+260. \$2.50.
- GATES, A. I. Interest and ability in reading. New York: Macmillan, 1930. Pp. xii+26+. \$1.50.
- HEPNER, H. W. Psychology in modern business. New York: Prentice-Hall, 1930. Pp. xv+728.
- HOLLINGWORTH, H. L. Social psychology: its concepts and theories. New York: Ronald Press, 1930. Pp. xi+590. \$4.50.
- KRUEGER, E. T., & RECKLESS, W. C. Social psychology. New York: Longmans, Green, 1931. Pp. viii+578.
- LUND, F. H. Emotions of men. New York: Whittlesley House (McGraw-Hill), 1931. Pp. x+348. \$2.50.
- MALLER, J. B. Cooperation and competition: an experimental study in motivation. *Teach. Coll. Contrib. Educ.*, 1930, No. 38+. Pp. x+176.
- MANUEL, H. T. The education of Mexican and Spanish-speaking children in Texas. Austin, Tex.: The Fund for Research in the Social Sciences, Univ. Texas, 1930. Pp. x+174. \$1.25.
- MAY, G. Social control of sex expression. New York: Morrow, 1930, 1931. Pp. xii+307. \$3.00.
- SOROKIN, P. A., ZIMMERMAN, C. C., & GALPIN, C. J. A systematic source book in rural sociology. Vol. 1. Minneapolis: Univ. Minn. Press, 1930. Pp. xx+645.
- WAHLIN, A. H. Livets princip: bevis för att livsprincipen följer elektrisk lagbundenhet. Stockholm: Author, 1930. Pp. 180.

THE INDIFFERENCE FUNCTION*

From the Department of Psychology, University of Chicago

L. L. THURSTONE

The purpose of this paper is to introduce a new problem in psychophysics which concerns also some fundamental economic theory. The problem in its economic setting is old, but the restatement of it in experimental form and its formulation as a psychophysical problem are probably new. The problem involves two psychological functions that we shall call the *satisfaction curve* and the *indifference curve* respectively. The general nature of these two functions will first be described, then the mathematical development of both functions from psychological postulates, and finally their experimental verification.

The formulation of this problem is due to numerous conversations about psychophysics with my friend Professor Henry Schultz of the University of Chicago. It was at his suggestion that experimental methods were applied to this problem in economic theory. According to Professor Schultz, it has probably never before been subjected to experimental study. The writer dares not venture far into the economic theory which may be implied in this psychophysical problem, but it is clear that here is a fertile field for investigation in a very old problem that overlaps economic theory and psychophysical experimentation.

If you have a certain amount of some desirable commodity it is not inconceivable that you would like to have some more. That is the kind of commodity we are studying here. We shall assume that your satisfaction from this commodity increases, the more you have of it, at least within the limits of plausible experimental or practical judgments. This fact we represent in Figure 1 in which s = satisfaction and x = number of items of a commodity. This rising satisfaction curve has been drawn so as to show an increase in satisfaction with increase in the amount of the commodity possessed by an individual. We are not concerned with those commodities which do not show such an increase in satisfaction, and of course we are not concerned with absurd extremes in quantity as, for example,

*Received in the Editorial Office July 22, 1930.

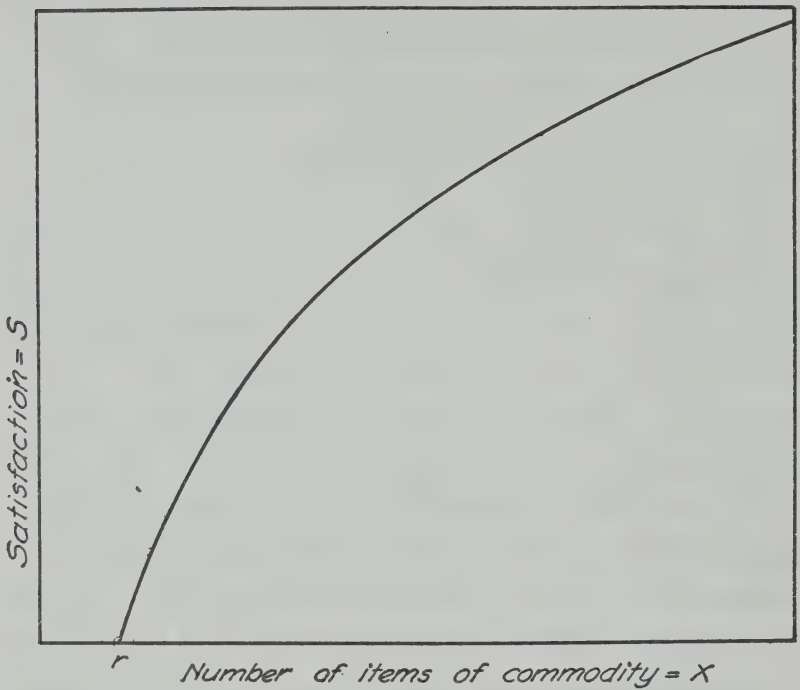


FIGURE 1

the immediate physical possession of a million apples for which our satisfaction might be obscured by embarrassment as to what to do with so many apples. Furthermore, Figure 1 is not concerned with the exchange or barter value of a commodity. For example, to possess a million apples would constitute wealth that could be exchanged for other things that we personally might desire more than a pile of apples, but we are concerned in this problem with the satisfaction that the owner may derive from a specified number of some commodity without regard to money cost. Our main question is rather to study the satisfaction that the owner derives from a specified number of the commodity itself.

It is frequently easy enough to count the number of items of a commodity that an individual possesses so that measurement of x constitutes no serious theoretical problem. In the measurement of satisfaction we shall imply throughout that it is accomplished in terms

of the subjective unit of measurement, the discriminial error, or some multiple of this unit (2, 3, 4).

In order to write an equation for the satisfaction curve we shall start with five fundamental psychological postulates, namely:

1. *Satisfaction increases with increase in the amount of the commodity possessed by an individual.* This assumption has been described above.

2. *There is a lower limit in the amount of the commodity below which the owner will not or cannot barter.* If, for example, you are accustomed to possess two new hats every year, this amount of the commodity might be so commonplace for you that it is taken for granted. You may regard it as a minimum necessity. Such a rate of consumption causes you neither noticeable pleasure nor pain. This lower limit in x is indicated in Figure 1 at r and the corresponding satisfaction is chosen as the origin for the scale of satisfaction. This postulate is not absolutely necessary for our problem, but it simplifies one aspect of it to designate the origin for s in this manner.

3. *Motivation is defined quantitatively as the anticipated increment in satisfaction per unit increase in the commodity.*¹ Here, as elsewhere in this problem, the satisfaction increments are expressed in terms of the discriminial error or as multiples of this unit. It is consistent with common sense that the motivation to acquire an additional unit of the commodity is smaller, the greater the amount already possessed. It is also clear that the motivation at any given value of x is the slope of the curve at that point.

Here we have defined the slope of the satisfaction curve at any particular point as motivation. It may be suggested that this definition of motivation as a quantitative psychological concept is by no means limited to the present problem. Stated more generally, motivation toward any specified goal, or toward unit accretion of any kind, may be defined as the increment in satisfaction represented by imaginal attainment and expressed in terms of the discriminial error. It is then possible to compare the motivation toward one object with the motivation of the same person toward some other object in a meaningful way and in quantitative terms.

4. *The motivation is finite when satisfaction is zero.* This simply means that, if you possess an amount of a commodity to which you are accustomed or which you take for granted, your motivation to-

¹This is said to be equivalent to the economist's "marginal utility."

ward an accretion cannot at that time be infinite. This assumes an origin for the continuum of satisfaction which facilitates a part of the solution, though not necessary for it, since the origin as defined can be treated as arbitrary.

5. *The motivation is inversely proportional to the amount already possessed.* This is our most fundamental psychological postulate. It is, in fact, the turning point in our rationalization of the problem which might proceed along several lines. This postulate is a simple one and it will be shown that it satisfies experimental results. This psychological postulate can be written more concisely in the form

$$\frac{ds}{dx} = \frac{k}{x} \quad [1]$$

in which

s = satisfaction, x = amount of the commodity possessed, $\frac{ds}{dx}$ = motivation, and k = a constant which characterizes the person and the particular commodity. Integrating, we have

$$\int ds = k \int \frac{dx}{x} \quad [2]$$

or

$$s = k \log x + c \quad [3]$$

which is certainly no stranger in psychophysics. It is our old friend, Fechner's law.

To those who are familiar with psychophysics it might seem as though the writer merely adopted Fechner's law in accordance with psychophysical habits, but such was not the case. As a matter of fact, many other psychological postulates have been tried instead of the fifth postulate above but the one which leads to Fechner's law seems to fit the experimental data better than any of the others that were tried. To the writer it still seems as though it would be psychologically preferable to start with the postulate that motivation is inversely proportional to the amount of *satisfaction* already attained from the commodity. That would lead to a square root law instead of the logarithmic law of Fechner. But in spite of much nursing of that hypothesis, the writer was not able to make the square root law fit the data so well as the law of Fechner. Therefore we returned to Fechner's logarithmic law which has been applied to the data in

this paper. There is still a possibility that the square root law will be found to be superior when the two hypotheses are compared with the same number of parameters so as to allow comparable degrees of freedom.

This is the equation represented in Figure 1 and it satisfies all of the psychological postulates listed above. The first postulate is satisfied by the fact that the curve of satisfaction rises with increase in amount of the commodity possessed. The second postulate is satisfied by placing the origin for satisfaction to correspond with some amount of the commodity such as r . It is clear that in the logarithmic equation the origin for satisfaction depends merely on the constant of integration and it can therefore be regarded as essentially in the nature of an arbitrary origin. Our experimental study does not include negative values for satisfaction. It would be possible to determine experimentally whether the law as stated in Equation 3 continues for negative values of S or whether it is necessary then to write a separate law for pain as we have done here for satisfaction. If such should be the case, then the origin would no longer be arbitrary but a functionally true origin. Since our experimental study does not cover negative values of S , we shall treat only the one logarithmic law for satisfaction as though it were continuous, and we shall regard the origin as essentially arbitrary in character.

The third postulate is really only a definition of motivation by which this psychological term may be given a quantitative formulation. It is at least consistent with the usual conversational meaning of the term motivation. The fourth postulate is satisfied by the equation in that the slope of the curve when satisfaction is zero cannot be infinite. In fact the slope of the curve is infinite only when satisfaction is $-\infty$. The fifth postulate which is the fundamental one is the simple differential equation which leads directly to Fechner's law.

Before using the satisfaction curve for the construction of the indifference curve, the general nature of the latter function will be described. Referring to Figure 2, let the variables x_1 and x_2 represent the amounts of two commodities that one person possesses. The point e in this figure represents the fact that one person owns a items of the first commodity and b items of the second. We shall suppose that both commodities are of such a character that he would like to have as many as possible of each of the two commodities. He might

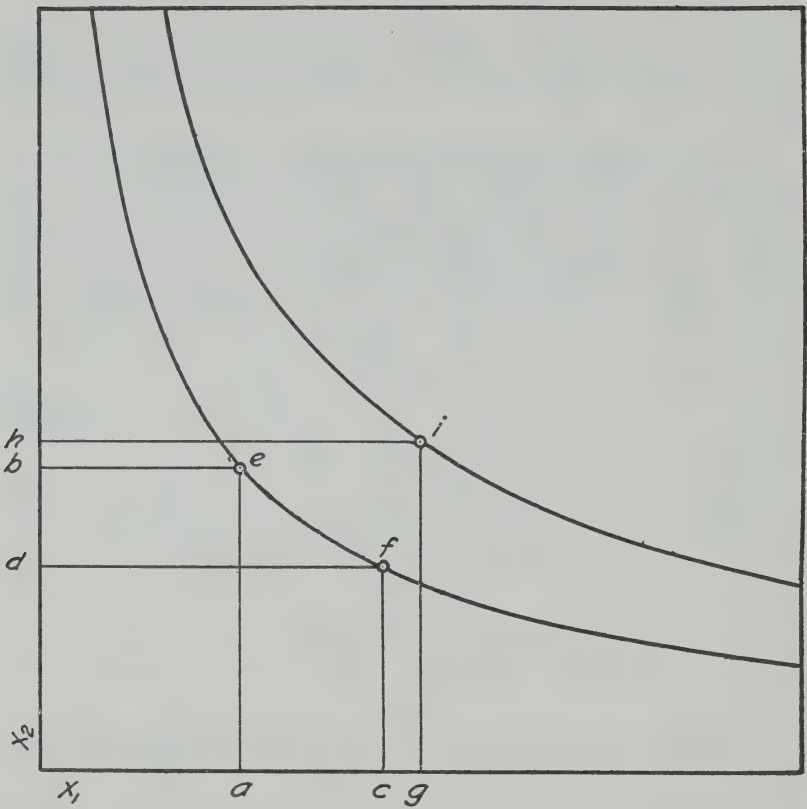


FIGURE 2

be willing to give up one item of the first commodity in order to obtain one or more items of the second. The rate at which he is willing to make the exchange depends of course on the actual quantities of the two commodities that he possesses at the time of barter, and it also depends on his relative preferences for the two commodities. Let us suppose that our subject is just barely willing to reduce his supply of No. 2 from b to d provided that he is given an increment in his supply of No. 1 from a to c . The latter situation in which he owns c items of No. 1 and d items of No. 2 is represented by the point f in Figure 2. Here we have two situations, e and f , which are equally attractive to the owner. If we imagine a curve drawn through all of the points of equal satisfaction, we shall have

an extended curve ef that we shall call an indifference curve. It may be so called because every point on such a curve represents a combination of quantities of the two commodities and the subject is indifferent as to which of the numerous combinations he possesses.

It is to be expected that such a curve will have a negative slope when we are dealing with commodities that are in general desirable in all the quantities that are practically feasible because a reduction in one of these commodities will be associated with an increment in the other one. It is as though a reduction in one of these quantities must be paid for by an increment in the other quantity. Needless to say, we are not here dealing with the money cost of these quantities. We are concerned with the satisfaction that they would give the owner irrespective of their money cost.

It is easy to imagine an indifference curve of positive slope to represent two quantities, one of which is in general desirable while the other one is regarded as undesirable. The subject would then judge whether he would be willing to accept so many items of a disadvantage in order to possess so many items that are desirable. The supposition is that as the amount of advantage is increased he would be willing to accept a greater amount of corresponding disadvantage. The slope of such an indifference curve would be positive, but we shall here limit ourselves to the comparison of commodities that are regarded as desirable in all of the amounts that can be ordinarily handled or imagined.

The total amount of satisfaction remains constant for all the combinations that are represented by points on the indifference curve. In moving from one point on this curve to any other point on the curve, we necessarily reduce the amount of one of the commodities and correspondingly raise the amount of the other commodity. The two simultaneous transactions constitute the barter by which satisfaction from one commodity is reduced by the same amount by which the satisfaction from the other commodity is raised. Hence we shall treat the problem as though the total amount of satisfaction were constant for all combinations represented by the indifference curve.

If we start with g items of the first commodity and h items of the second, then, by Figure 2, we shall have a greater amount of total satisfaction than at either e or f because we are starting with a larger amount of each of the two commodities. It is clear that another indifference curve may be drawn through the point i in the same manner. This makes it evident that in Figure 2 we might draw a

family of indifference curves such that the total satisfaction remains constant throughout any one of the indifference curves. The several curves would differ in the total amount of satisfaction that they represent. For example, the curve through the point *i* represents a higher total amount of satisfaction than the curve *ef*. We can look upon these indifference curves as contour lines on a topographic map. If a solid model were constructed with Figure 2 as a base, the vertical dimension would represent total satisfaction which would of course be high near the upper right corner of the diagram. The lowest point in the model would be at the lower left corner of the diagram. In the topographic analogy Figure 2 would represent a hill with its top in the upper right corner of the diagram and with the indifference curves representing points of constant elevation or total satisfaction. (See Figure 18.)

A rather simple objection might be raised in the assertion that the kind of satisfaction derived from one of these commodities might be entirely different from the kind of satisfaction derived from the second. This is true, but it is also true that all of us make decisions daily of the very kind that is implied by the indifference curves. For example, if you cannot buy a radio and a fur coat at the same time, you must decide whether to have one or the other or something else, such as a bank balance, even though the satisfactions involved in these various purchases are qualitatively quite different. We shall assume that for analytical purposes any purchase or barter means the reduction in satisfaction from one commodity and an equal increment in satisfaction from the other. In practice it is probable that the positive increment must at least slightly exceed the negative decrement in order to effect barter.

We shall now make use of the equation of satisfaction in order to write the equation of the indifference curve. We may rewrite Equation 1 specifically for one of the two commodities which is indicated by the subscript.

$$\frac{ds_1}{dx_1} = \frac{k_1}{x_1} \quad [4]$$

The same equation may be written for the second commodity as follows:

$$\frac{ds_2}{dx_2} = \frac{k_2}{x_2} \quad [5]$$

From [5] it is clear that

$$\frac{dx_2}{ds_2} = \frac{x_2}{k_2} \quad [6]$$

The indifference curve is drawn so that an increment in satisfaction from one of the commodities in simple barter is, in absolute magnitude, equal to the decrement in satisfaction for the other commodity and hence

$$ds_1 = -ds_2 \quad [7]$$

Therefore, from Equations 4, 6, and 7, we may write

$$\frac{dx_2}{ds_2} \cdot \frac{ds_1}{dx_1} = -\frac{dx_2}{dx_1} = \frac{k_1 \cdot x_2}{k_2 \cdot x_1} \quad [8]$$

Integrating, we have

$$k_2 \int \frac{dx_2}{x_2} = -k_1 \int \frac{dx_1}{x_1} \quad [9]$$

or

$$k_1 \log x_1 + k_2 \log x_2 = \log m \quad [10]$$

in which k_1 , k_2 , and $\log m$ are three constants. The two constants k_1 and k_2 represent the different rates at which satisfaction increases with increase in the amounts of the two commodities respectively. It will be shown that the constant $\log m$ represents essentially the elevation or total amount of satisfaction represented by the indifference curve.

The above equation may be written in exponential form as follows:

$$x_1^{k_1} \cdot x_2^{k_2} = m \quad [11]$$

and this may be regarded as the general equation of the indifference function to which we have arrived by assuming that Fechner's law is applicable to the satisfaction curve. This is synonymous with the assumption that motivation is inversely proportional to the amount of each commodity already possessed.

The constant $\log m$ of the indifference equation can be defined further as follows. If we rewrite Equation 3 specifically for each of the two commodities, we have

$$s_1 = k_1 \log x_1 + c_1 \quad [12]$$

$$s_2 = k_2 \log x_2 + c_2 \quad [13]$$

and hence

$$s_1 + s_2 - c_1 - c_2 = k_1 \log x_1 + k_2 \log x_2 \quad [14]$$

But from Equation 10 it follows that

$$s_1 + s_2 - c_1 - c_2 = \log m \quad [15]$$

From this equation it is clear that

$$s_1 + s_2 = \text{a constant} \quad [16]$$

which is what we should expect from our general analysis of the nature of simple barter.

The interpretation of Equations 10 and 11 may be facilitated by means of Figure 3 in which two satisfaction curves are represented on the same diagram. Let the two curves, 1 and 2, be the satisfaction curves for two commodities. They are drawn so as to represent different values for the constants k_1 and k_2 and they are also so drawn that the values of r_1 and r_2 are different. Let us start with the possession of a items of the first commodity and a' items of the

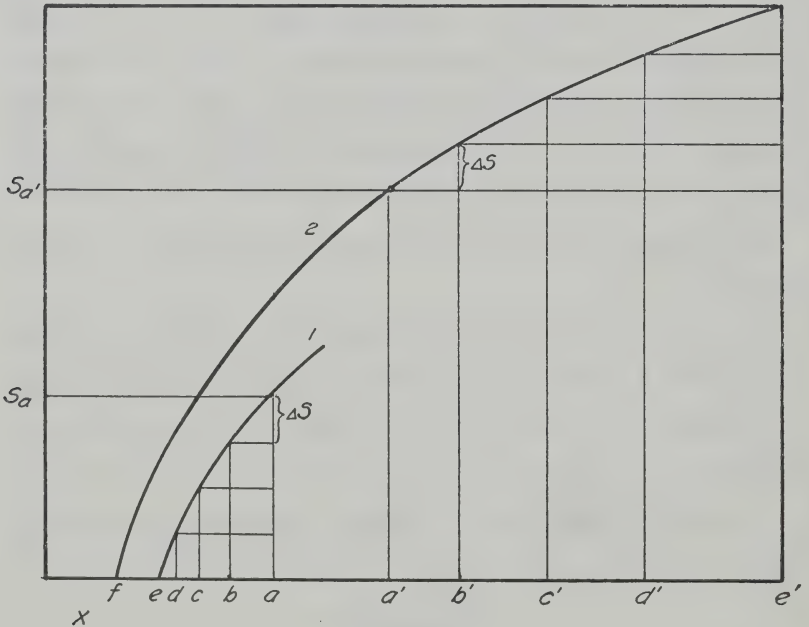


FIGURE 3

second commodity. The corresponding satisfactions are represented by the two ordinates S_a and $S_{a'}$ respectively. The possession of both of these quantities is assumed to give the owner a total satisfaction equal to the sum of these two ordinates, namely, $(S_a + S_{a'})$ and the combination a and a' can of course be represented as a point on an indifference curve.

Suppose that the subject is indifferent about a proposal to reduce the quantity a to the quantity b at the same time that the quantity a' is augmented to the quantity b' . Then the satisfaction S_a suffers a decrement of ΔS while the satisfaction $S_{a'}$ is augmented by the same amount, namely, ΔS . This trade is represented graphically. In this manner it will be possible to reduce S_a until it vanishes. In the diagram this is shown in four successive equal decrements, ΔS , by which the quantity a is reduced successively to the values b, c, d , and e . At the same time the satisfaction $S_{a'}$ is augmented by four successive identical increments, ΔS , by which the quantity a' is augmented successively to the values b', c', d' , and e' . It is clear therefore that the five points aa', bb', cc', dd' , and ee' , all lie on the same indifference curve because the total satisfaction derived from the two commodities is constant. Every decrement in the satisfaction from one of the commodities is exactly balanced by an equal increment in the satisfaction from the second commodity.

We shall interpret Figure 3 to mean that the subject is willing to barter by reduction in the quantity a for increases in the quantity a' as long as he has any satisfaction from the first commodity to give up, but that there is a lower limit below which he will not barter. This lower limit is e items of the first commodity and f items of the second commodity. In a practical situation it means that we might be willing to get along with fewer shoes in order to have a large supply of new hats but that there is a limit below which we will not barter away our shoes for any number of hats. For such commodities as hats and shoes these limits are determined partly by physical necessity and partly by the economic or social level at which the subject is living. These quantities e and f are interpreted as more or less taken for granted by the subject so that any further reduction in either of them would be regarded with actual displeasure or pain which cannot be offset by any quantity of a different commodity.

In Figure 4 we have drawn three satisfaction curves with different

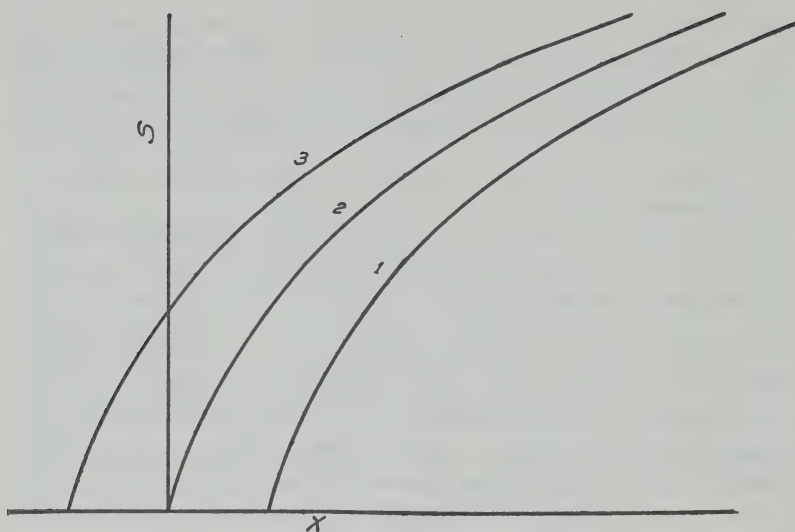


FIGURE 4

x -intercepts. Curve 1 represents the type of function that we have used for the above illustrations. But it is possible of course to have a satisfaction curve which passes through the origin in the sense that it causes neither pleasure nor pain to be without the particular commodity. Curve 2 is so drawn, however, as to represent an increase in satisfaction from ownership of that commodity, while its absence does not result in actual pain or displeasure. It is essential to distinguish between motivation and satisfaction. By motivation we mean the strength of the urge toward an accretion. It has been defined as the amount of satisfaction that is represented by an imaginal unit accretion. The satisfaction, however, represents the summation of positive affect that has already been derived from the commodity in question. It is therefore possible to have none of a commodity and to have experienced no satisfaction from it and yet have some motivation toward acquiring it.

Curve 3 in Figure 4 represents the situation in which the subject has experienced some satisfaction from a commodity without owning any of it. This is of course possible, and it is also conceivable that the motivation or slope of the curve at the s -axis is therefore smaller than it would be if the same subject did not experience any satisfaction from the commodity at all.

EXPERIMENTAL PROCEDURE

Having developed rational equations for the satisfaction curve and for the indifference curve, our next step is to ascertain whether they can be verified experimentally. In Figure 5 we have represented hats and shoes as the two commodities. Any point on this diagram represents a combination such as eight hats and eight pairs of shoes. If the curve in Figure 5 were an indifference function, then the subject would say that eight hats and eight pairs of shoes would give him as much satisfaction as six hats and ten pairs of shoes because both of these points lie on the curve. Perhaps the simplest experimental method that comes to mind is to ask a subject to fill in the blank space in a series of choices of the following type:

eight hats and eight pairs of shoes
or six hats and ——— pairs of shoes

If the subject would fill in the blank space with the number of pairs of shoes which would give him as much satisfaction to own as the first combination of eight and eight, and if similar judgments were made for a series of quantities of hats, more than six as well as less than six, it would of course be possible to plot the indifference curve. This is the method of reproduction as it might be applied to this problem. In the present situation the judgments would probably be so unstable and so markedly influenced by the desire for numerical consistency that the curve so obtained would be of doubtful value. It is possible that the method of reproduction could be applied to this problem in a manner which would make it experimentally satisfactory, but in the present study we have used the constant method.

The constant method takes the following form. One of the combinations such as eight hats and eight pairs of shoes is chosen as a standard and each of the other combinations is compared directly with it. Thus in Figure 5 we should expect to find that if the subject were asked to choose (eight hats and eight pairs of shoes) or (seven hats and fourteen pairs of shoes) he might be quite willing to give up one of the hats in order to possess six additional pairs of shoes, assuming of course that the money cost to him were the same. We should therefore expect to find the point *B* marked plus because the combination at that point is preferred to that of the standard at *A*. This may be judged from Figure 5 because the point *B* lies on an indifference curve at a higher elevation of satisfaction than the point *A*.

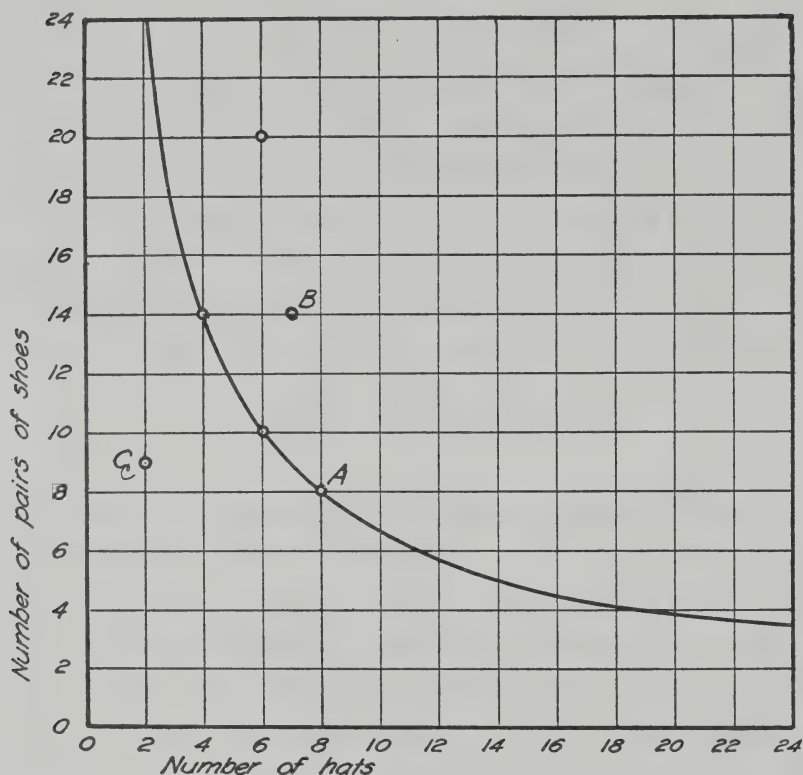


FIGURE 5

On the other hand, if the subject were asked to choose either the combination at *A* or the one at *C*, he would probably prefer *A*, since very likely he would not care to give up six of his eight hats in order to get only one additional pair of shoes. We should expect the point *C* to be marked with a minus sign to show that the subject preferred the standard at *A*. If all of the judgments were consistent we should expect to find only plus signs above the indifference curve and only minus signs below the curve. In actually drawing the indifference curve we record the preferences of the subject in the form of plus and minus signs in a diagram and we then draw the indifference curve so that, as far as possible, there are only plus signs above the curve and only minus signs below it. Naturally, the records from only one subject who makes only one judgment for each combination cannot be ex-

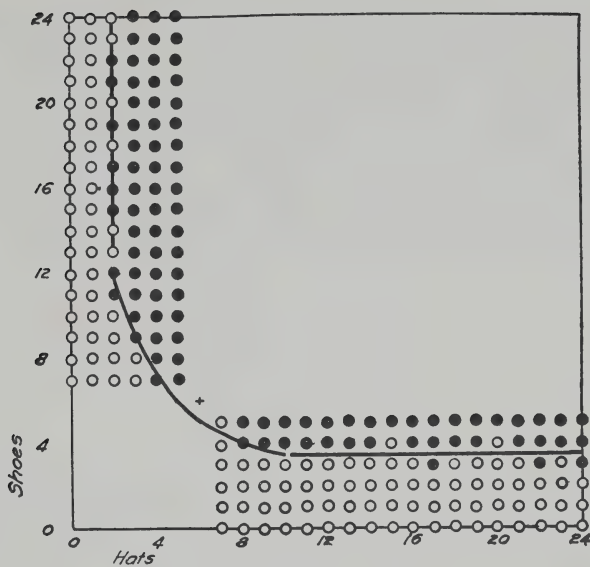


FIGURE 6

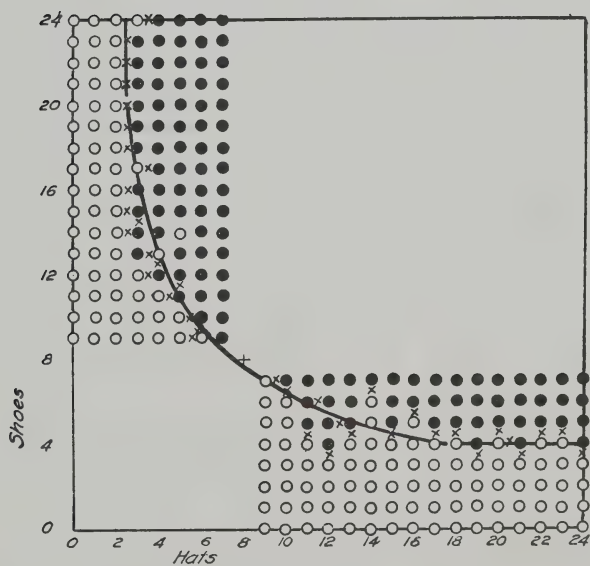


FIGURE 7

pected to be free from inversions, especially in the field close to the curve where the difference in satisfaction is at best rather slight. If our analysis of the function is correct, it should be possible to draw a curve of Equation 10 so as to satisfy the above described conditions.

In order to make the test of Equation 10 fairly complete we have carried out the test with four standards for each of the three possible pairs of three commodities. The procedure was simply to present to the subject a list of alternatives of the following type.

8 hats and	8 shoes
6 hats and	9 shoes
8 hats and	8 shoes
4 hats and	15 shoes
8 hats and	8 shoes
9 hats and	3 shoes

The subject was asked to indicate for each alternative which of the two combinations would probably give him the more satisfaction on the assumption that the two alternative combinations would cost him the same and with the further assumption that the articles would be of his free selection within the general range in price and quality to which he is accustomed. If the standard was checked, the corresponding variable combination was marked with a minus sign on the diagram. If the variable were checked that variable combination was checked with a plus sign on the diagram. Evidently it would be useless to ask a subject to choose between eight hats and eight shoes or ten hats and ten shoes because in such a presentation both of the commodities are augmented in quantity and the obvious answer is of course to check the greater quantity of both commodities. For this reason we limited ourselves to those alternatives in which one of the commodities is reduced while the other is augmented. In this manner it required an act of judgment beyond the mere inspection of the numbers to make an intelligent choice. These judgments are sufficiently numerous to allow a fair determination of the course of the indifference curve.

The subject whose records are here analyzed was entirely naïve as regards the psychophysical problem involved and had no knowledge whatever of the nature of the curves that we expected to find. The judgments were made as indicated above in random order, and the subject did not make any tabulation or analysis of the judgments.

In Figure 7 we have a record of the actual judgments made by

our subject when the standard combination was eight hats and eight pairs of shoes. Every small black circle indicates a judgment in which the variable combination was preferred, while every small open circle represents a judgment in which the standard was preferred. For example, in Figure 7 we find that our subject would rather have six hats and fourteen pairs of shoes than to have eight of each. In a similar manner she would rather have eight hats and eight pairs of shoes than to have twelve hats and three pairs of shoes. All of the other black circles and open circles are to be interpreted in the same manner.

If the judgments were perfectly consistent, it would be possible to draw the indifference function so that all of the black circles lie above the curve and all of the open circles lie below the curve. This would mean that all of the combinations represented by points above the curve represent greater satisfaction than the combinations represented by the points that lie on the curve. Similarly, all points below the curve would represent combinations that are less desirable than those that lie on the curve. Since every circle in these diagrams represents only one judgment by a single subject, it is to be expected that there will be inversions, especially in view of the fact that the difference in satisfaction represented by neighboring points is not very marked. Our problem is to draw the best-fitting indifference curve through the field of black and white circles so that the black circles lie, as far as possible, above the curve, while the open circles lie below it.

It is evident that the black and white circles represent two fields such that the ideal indifference curve should really separate them. In order to locate a series of points for the purpose of curve fitting we proceed as follows. In Figure 7 we note that the column for four hats has twelve black circles and four white ones. We put a small cross as shown in the figure at such a point in this column that there are as many inversions above it as there are inversions below it. In this column it happens that there is one white circle above the cross and one black circle below it. We proceed in this manner for every column and for every row in each of our diagrams. In the column for three hats, for example, the small cross is so located that there are two white circles above it and two black circles below it. In the horizontal row for six shoes the small cross is so located that there is one black circle to the left of the cross and one white circle to the right of it. All of these crosses are located

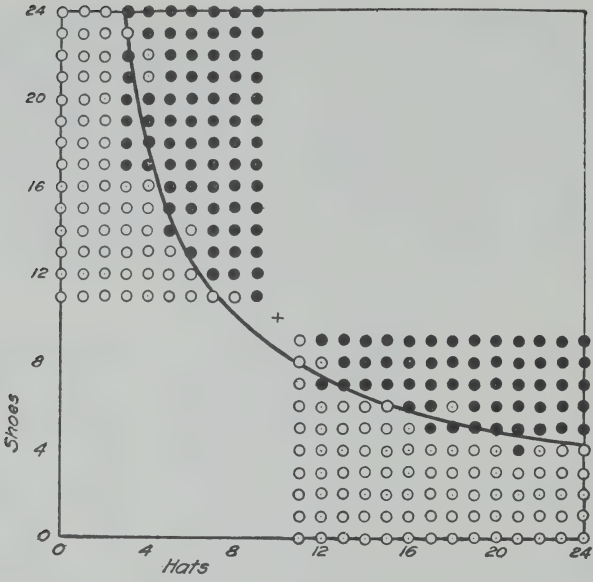


FIGURE 8

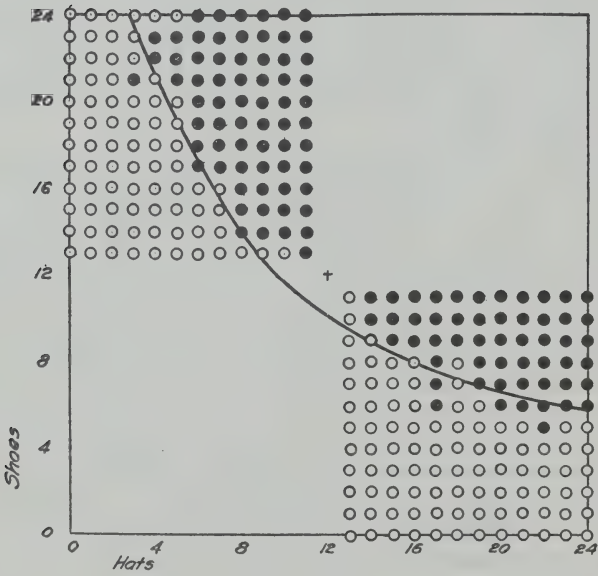


FIGURE 9

in this manner in Figure 7. Usually there are either no inversions or only one inversion. Occasionally there are two inversions.

Equation 10 is to be fitted by the values represented by the small crosses in Figure 7. We write the equation in summation form as follows:

$$k_1 \sum \log x_1 + k_2 \sum \log x_2 = n \log m$$

In order to define the s -scale we let the constant k_1 be unity. Then there are two consonants to be determined, namely k_2 and $\log m$. We determine these two constants by the simple method of averages which need not be described in detail here. The constants for Figure 7 are $k_2 = 1.39$ and $\log m = 1.80$. In Table 1 we have summarized the constants for each of the figures from 6 to 13 inclusive. These are the comparisons of hats with shoes, and of hats with overcoats.

Inspection of Figures 6-9 inclusive shows that Figure 6 is a curve much closer to the coordinates than the other three. It is probable that the judgments in the vicinity of the curve in Figure 6 are not so trustworthy as the judgments in the other three figures, partly because of the fact that the curve in Figure 6 is rather close to the level of satisfaction at which the subject refuses to barter away the small quantities of either commodity. For this reason we have chosen the average value of k_2 from Figures 7-9 inclusive as the general value for this constant, namely, 1.26. In a similar manner, the average value for k_3 is 1.32. The values of the constants k_1 , k_2 , and k_3 are summarized at the bottom of Table 1. These values indicate that for our subject the satisfaction curve rises fastest for overcoats and slowest for hats. In other words, the increment of satisfaction per unit accretion of overcoats is larger than the satisfaction derived

TABLE 1

Figure	k_2/k_1	k_3/k_1
6	1.39	
7	1.30	
8	1.28	
9	1.19	
10		1.57
11		1.25
12		1.34
13		1.37
<hr/>		
	$k_1 = 1.00$	
	$k_2 = 1.26$	
	$k_3 = 1.32$	

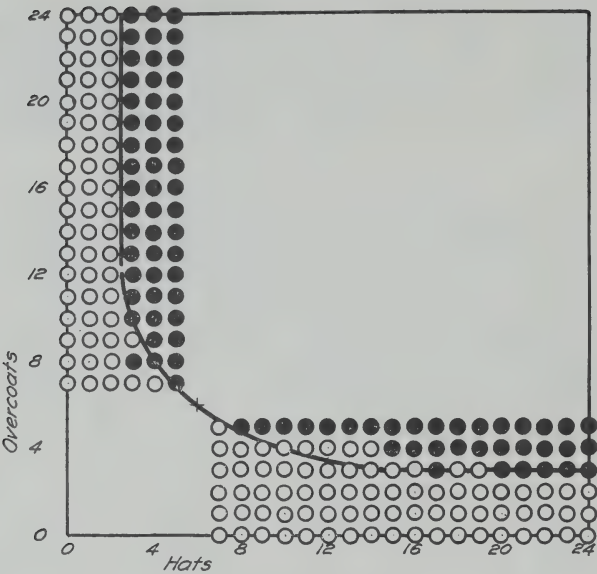


FIGURE 10

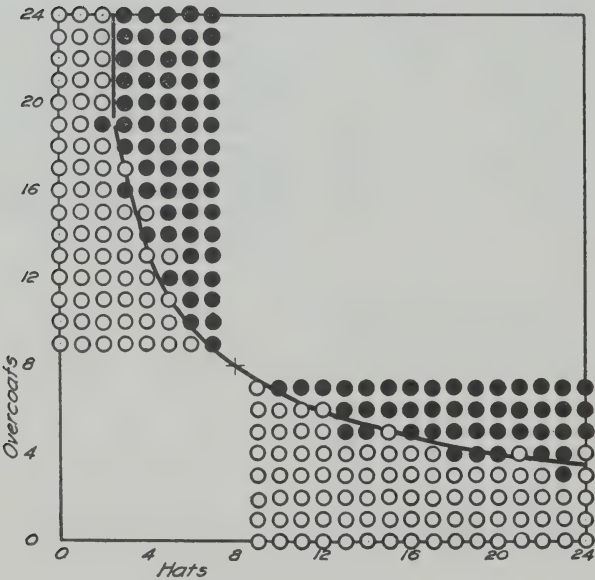


FIGURE 11

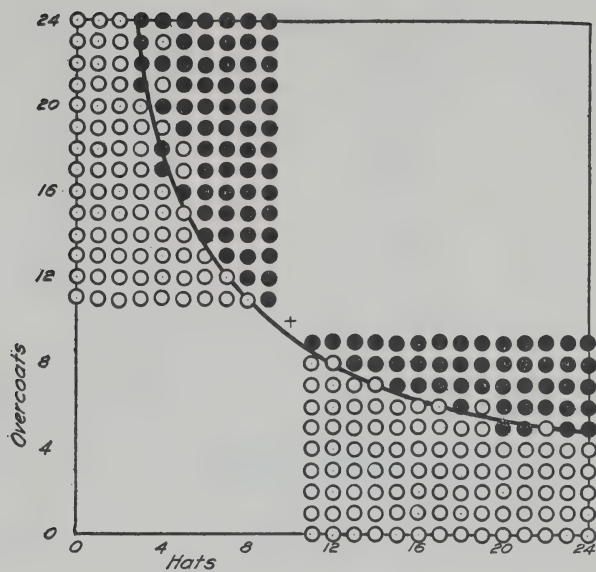


FIGURE 12

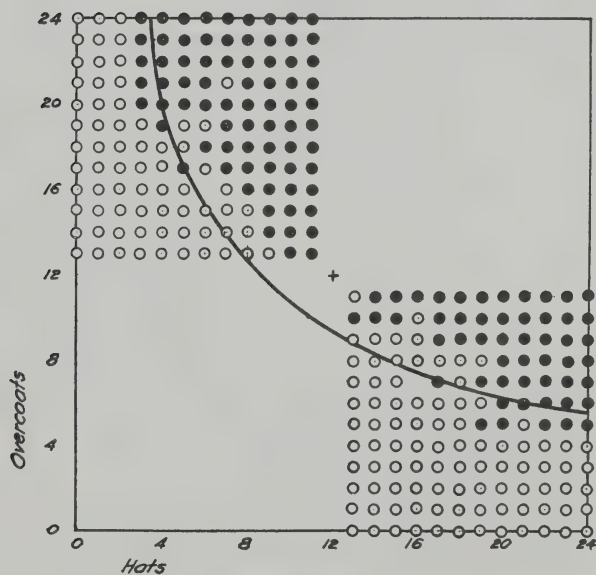


FIGURE 13

TABLE 2

Figure	$\log m$	$\log m_s$	e
6	1.67	1.76	— .09
7	2.01	2.04	— .03
8	2.14	2.26	— .12
9	2.34	2.44	— .10
10	1.79	1.81	— .02
11	2.11	2.10	+ .01
12	2.28	2.32	— .04
13	2.35	2.50	— .15

$$E = \text{Constant error} = \text{average } e = -.066$$

for each additional hat when these increments are measured from the same satisfaction level. This is perhaps what one might expect from the common-sense values of hats and overcoats. It is also apparent, that, for our subject, additional shoes give larger increments of satisfaction than is derived from additional hats.

Having determined the best-fitting values of k_2 and k_3 , we can now determine the value of $\log m$ for each diagram. This is done again by the method of averages and the values are listed in the second column of Table 2. These values would be altered slightly by different methods of calculation. The satisfaction $\log m$, which is represented by the standard combination for each curve, can be ascertained from the same equation in which x_1 and x_2 are the coordinates of the standard combination. Let this be designated $\log m_s$, and let the corresponding constant for the equation of the best fitting curve be designated $\log m_a$. The difference between these two constants is the constant error of the method of right and wrong cases. It is the difference in elevation between the indifference curve that passes through the standard combination and the best-fitting indifference curve through the small crosses that were drawn in each diagram. The constant error for each figure is listed in the last column of Table 2. The constant error for the method of right and wrong cases is designated E and it is taken to be the average of the constant errors for the eight diagrams. The value of E , which is the average value of the eight values of e , is $-.066$.

Now that we know the values of k_1 , k_2 , and k_3 , and the constant error of the psychophysical method involved, which is $-.066$, we should be able to predict the indifference curves for the comparison of shoes and overcoats. So far we have made use of only two sets of

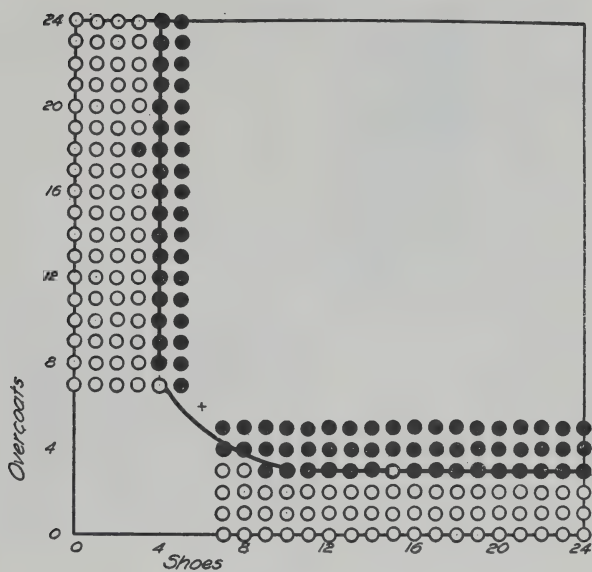


FIGURE 14

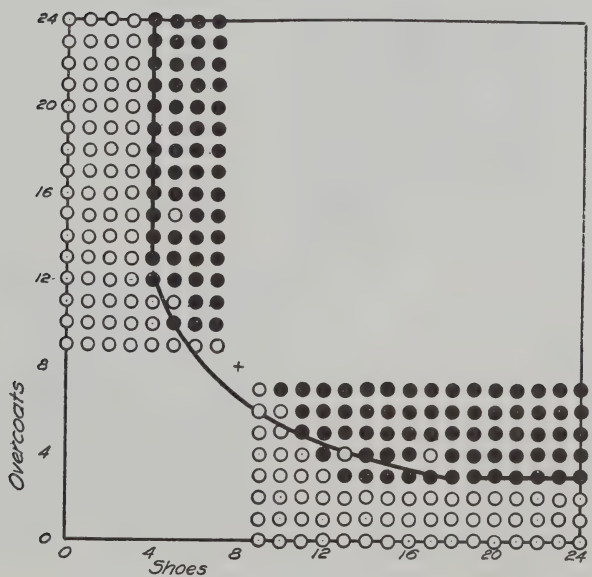


FIGURE 15

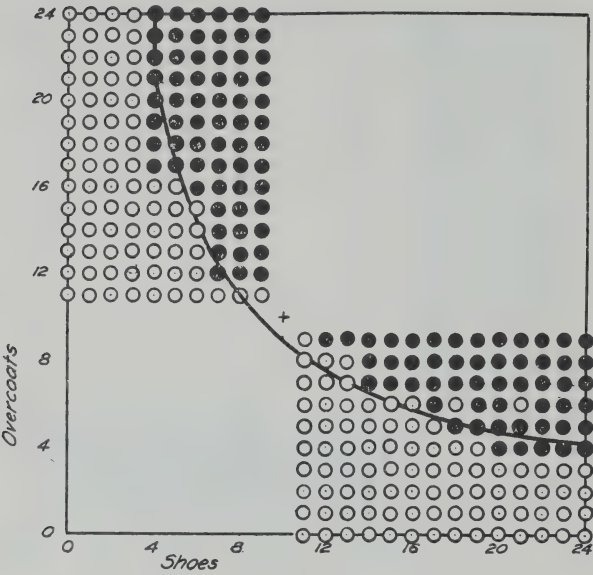


FIGURE 16

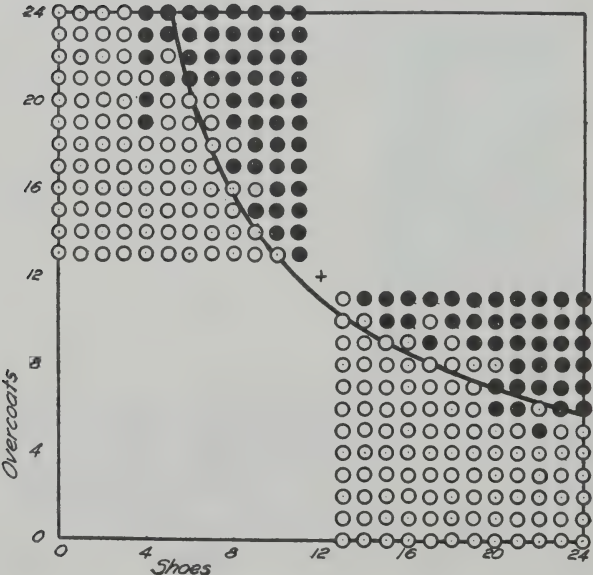


FIGURE 17

TABLE 3

Figure	x_2	x_3	$\log m_s$	$\log m_a$
14	6	6	2.01	1.94
15	8	8	2.33	2.26
16	10	10	2.58	2.51
17	12	12	2.78	2.72

comparisons, namely hats-shoes, and hats-overcoats. We shall now predict the equations of the indifference curves for the third comparison, namely, shoes-overcoats. Since we have the experimental data for all three sets of comparisons, we can ascertain how closely the third set of indifference curves can be predicted from the known constants, derived from the first two sets of comparisons. This constitutes the test of the fundamental psychological hypothesis that is involved, namely, that the satisfactions from several commodities are summative when all of the quantities involved are above the level which the subject regards as the level of absolute necessity.

In Table 3 we have listed the four standards for the third set of comparisons, namely, shoes-overcoats. These comparisons are shown graphically in Figures 14-17 inclusive. The standard combinations are listed in the second and third columns. The value of $\log m_s$ is listed in the next column. It is calculated again by Equation 10 with the known values of k . But the best-fitting curves for the first two sets of comparisons revealed a slight constant error in $\log m$. Applying the same constant error to the predicted third set of comparisons, we have the predicted values of $\log m_a$ which are listed in the last column of Table 3. With these values we can write the four predicted equations and plot each of them.

Predicted equations for the indifference curves (shoes-overcoats) with four standard combinations	Standard combination	
	Shoes	Overcoats
$1.26 \log x_2 + 1.32 \log x_3 = 1.94$	6	6
$1.26 \log x_2 + 1.32 \log x_3 = 2.26$	8	8
$1.26 \log x_2 + 1.32 \log x_3 = 2.51$	10	10
$1.26 \log x_2 + 1.32 \log x_3 = 2.72$	12	12

Inspection of Figures 14-17 inclusive indicates that the agreement between the predicted curves and the distributions of black and white circles is quite satisfactory. In general, the black circles lie

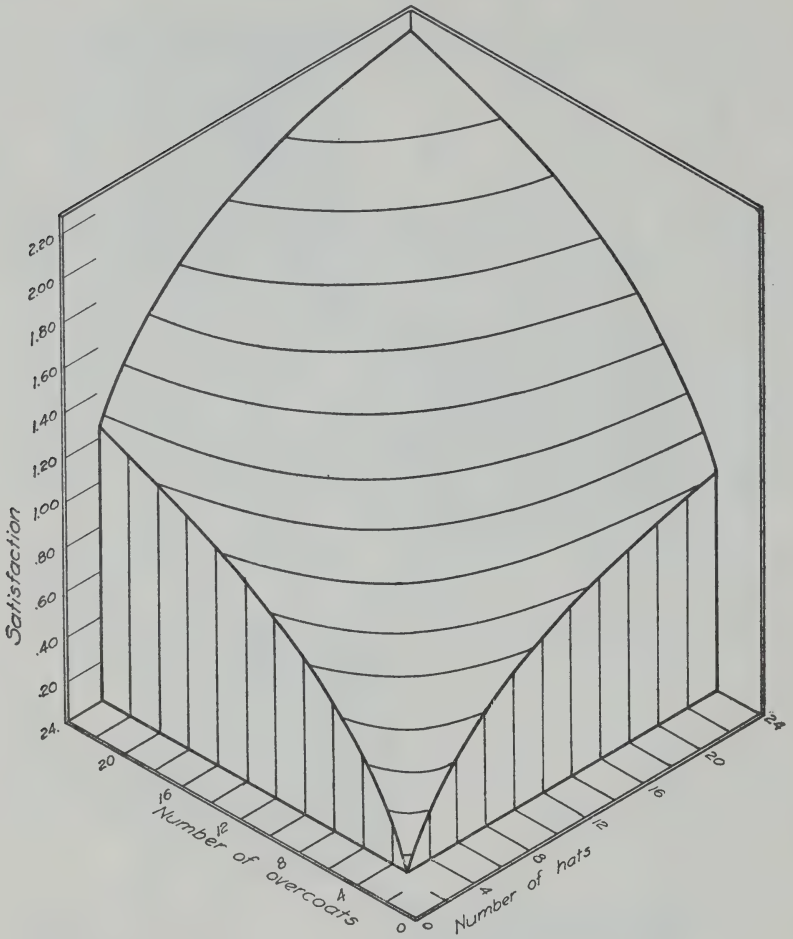


FIGURE 18

above the predicted indifference curve, while the white circles lie below it.

In each of the indifference curves that run close to the two coordinates it is apparent that there is a lower limit for each commodity below which the subject is not willing to barter. According to Equation 10, the indifference curve should be asymptotic to the two coordinates. But, as a matter of fact, the subject refuses to barter

for certain minimum quantities of each commodity. When the number of hats is lower than three the subject does not barter away any of them no matter how many shoes or overcoats are offered in exchange. The lower limit for shoes is four pairs, and the lower limit for overcoats is three. The curves in our diagrams are drawn to these lower limits but not beyond them. It is clear, therefore, that while the indifference function should be asymptotic to the coordinates according to the equation, the experimental function does not extend beyond the lower limits indicated. This fact can be interpreted to mean that the commodities have certain values for exchange purposes which depend on the quantities possessed, but that all barter ceases entirely when the quantity of a commodity reaches a lower limit which the subject evidently regards as a necessity.

CONCLUSIONS

The object of this paper has been to show that it is possible to reduce the indifference function to experimental treatment and that it is possible to write a rational equation for the indifference function which is based on plausible psychological postulates. We have developed two psychological functions, namely, the curve of satisfaction and the indifference curve. The equation for the satisfaction curve which agrees with the experimental data is similar in form to Fechner's law, and it may, in fact, be regarded as an extension of that law which has been ordinarily interpreted as limited to sensory discrimination. The rational equation for the indifference curve takes

the general form $x_1^{k_1} \cdot x_2^{k_2} = m$, in which x_1 and x_2 are the quantities of the two commodities. The constants k_1 and k_2 are descriptive of the individual subject and his preferences for the two commodities, while the constant m designates the total amount of satisfaction which is represented by the indifference curve. This equation is based on the fundamental psychological postulate that motivation toward accretion in each commodity is inversely proportional to the amount of the commodity already possessed. We have also offered a quantitative definition of motivation as the amount of anticipated satisfaction per unit increase in the commodity. This definition of the concept of motivation makes it possible to treat motivation generally in a quantitative manner and it lends itself to other psychological problems that involve motivation. Throughout this paper it has been assumed that increments in satisfaction are

measured in terms of the psychological unit of measurement, the discriminial error, or multiples of that unit. We have described the experimental records for only one subject. The same procedures have been tried on several other subjects with similar results.

REFERENCES

1. FISHER,—. Mathematical investigations in the theory of values and prices. New Haven, Conn.: Yale Univ. Press.
2. THURSTONE, L. L. Psychophysical analysis. *Amer. J. Psychol.*, 1927, **38**, 368-389.
3. ————. A mental unit of measurement. *Psychol. Rev.*, 1927, **34**, 415-423.
4. ————. The measurement of psychological value. In *Essays in philosophy*. Chicago: Open Court Publ. Co. Pp. 157-174.

University of Chicago
Chicago, Illinois

LA FONCTION DE L'INDIFFÉRENCE

(Résumé)

L'objet de cet article est de montrer qu'il est possible de réduire la fonction de l'indifférence au traitement expérimental et qu'il est possible d'écrire une équation rationnelle pour la fonction de l'indifférence laquelle est basée sur des postulats psychologiques plausibles. On a développé deux fonctions psychologiques, c'est-à-dire, la courbe de satisfaction et la courbe d'indifférence. L'équation pour la courbe de satisfaction laquelle s'accorde avec les données expérimentales est semblable en forme à la loi de Fechner, et on peut, en effet, la considérer comme une extension de cette loi qu'on a interprétée ordinairement comme limitée à la discrimination sensorielle. L'équation rationnelle pour la courbe d'indifférence prend la forme générale $x_1^{k_1} \cdot x_2^{k_2} = m$ où x_1 et x_2 sont les quantités des deux choses. Les constants k_1 et k_2 décrivent le sujet individuel et ses préférences pour les deux choses, tandis que le constant m désigne la somme totale de satisfaction représentée par la courbe d'indifférence. Cette équation est basée sur le postulat psychologique fondamental que la motivation vers un accroissement de chaque chose est inversement proportionnelle à la somme de la chose déjà possédée. On a offert aussi une définition quantitative de la motivation comme la somme de satisfaction anticipée pour chaque accroissement dans la chose. Cette définition du concept de motivation le rend possible de traiter la motivation généralement d'une façon quantitative et elle se prête à d'autres problèmes psychologiques où il s'agit de la motivation. Toujours dans cet article on suppose que les accroissements de satisfaction sont mesurés en termes de l'unité psychologique de mesure, l'erreur de discrimination, ou multiples de cet unité. On a décrit les résultats expérimentaux seulement pour un sujet. On a employé les mêmes processus sur plusieurs autres sujets avec des résultats pareils.

THURSTONE

DIE GLEICHGÜLTIGKEITSFUNKTION

(Referat)

Es war das Ziel dieser Abhandlung, zu erweisen, dass es möglich ist, die Gleichgültigkeitsfunktion (indifference function) experimentell zu behandeln, und das man für die Gleichgültigkeitsfunktion eine rationelle Gleichung schreiben kann, welche auf plausiblen psychologischen Voraussetzungen gegründet ist. Wir haben zwei psychologische Funktionen entwickelt, —namentlich die Befriedigungskurve (curve of satisfaction) und die Gleichgültigkeitskurve. Die Gleichung für die Befriedigungskurve welche mit den experimentellen Daten übereinstimmt gleicht in ihrer Form dem Fechnerschen Gesetz, und man kann sie eigentlich als eine Ausdehnung dieses Gesetzes betrachten, welch letzteres gewöhnlich als auf sensorielle Unterscheidung (sensory discrimination) beschränkt gedeutet worden ist. Die rationelle Gleichung für die Gleichgültigkeitskurve nimmt folgende allgemeine Form an: $x_1^{k_1} \cdot x_2^{k_2} = m$, worin x_1 und x_2 die Grössen (quantities) der zwei Artikeln (commodities) darstellen. Die Konstanten k_1 und k_2 beschreiben die einzelne Versuchsperson und seine Vorlieben für die zwei Sachen, während die konstante Quantität m den Gesamtbetrag an Befriedigung darstellt, der durch die Gleichgültigkeitskurve angedeutet wird. Diese Gleichung stützt sich auf die fundamentale psychologische Voraussetzung, dass der Antrieb (motivation) nach Zuwachs an jedem Gegenstand desto stärker ist, je weniger man schon von dem Gegenstand besitzt. Wir haben auch den Antrieb begriffsbestimmend quantitativ geschildert als den Betrag an erwarteter Befriedigung per Einheit Zuwachs an dem Gegenstand (per unit of increase in the commodity). Diese Begriffsbestimmung des Begriffes "Antrieb" ermöglicht die Behandlung des Antriebs im Allgemeinen auf quantitativer Weise, und es lässt sich bei anderen psychologischen Fragen anwenden, bei denen der Antrieb in Anspruch kommt. Durchaus in dieser Abhandlung ist angenommen worden, dass Zunahmen an Befriedigung gemessen werden können mit Bezug auf die psychologische Masseinheit (unit of measurement), den Unterscheidungsfehler, oder Vervielfachungen dieser Einheit. Wir haben die experimentellen Ergebnisse über bloss eine Versuchsperson beschrieben. Die selben Verfahren sind an mehreren anderen Vpn. mit ähnlichem Erfolg probiert worden.

THURSTONE

A STUDY OF THE OPINIONS OF COLLEGE STUDENTS*

From the North Carolina State College

K. C. GARRISON AND MARGARET MANN

Opinion deals with that which is probable and subject to change. It rests in the main upon the sentiment of the individual. Thurstone (8) has defined it as "the verbal expression of an attitude"; and an attitude as "the sum total of man's inclinations and feelings, prejudice or bias, preconceived notions, fears, ideas, threats, and convictions about any specific topic." The study and measurement of attitudes is of such recent development that no method for its prosecution has as yet been formulated.

Opinion reflects the early environment of the students, together with a certain amount of reading with deliberations. A review of all the literature pertaining to opinions or beliefs would not be appropriate to a scientific attempt to investigate and analyze the beliefs of college students. Hence only those investigations of recent date and scientific in nature, dealing specifically with the measurement and analysis of beliefs of students, are included in this review. Case studies, historical treatises, personality and character trait studies as well as specialized studies of the attitudes of groups other than students are all of interest and value but not pertinent to this study.

In a fairly recent investigation conducted at the University of Washington, Guthrie (3) obtained from 5 to 126 opinions of judgments on over 100 teachers from more than 800 students. In repeating his experiment, he found a correlation of .89 with his original returns and concluded that student opinion was comparatively reliable from one estimate to another.

That college training has a very slight effect upon the real opinions of students was a most significant general educational implication of a study conducted by Jones (4) at the University of Buffalo, in which a questionnaire was submitted to 418 students. He found the seniors more conservative in regard to economic and more liberal on religious questions. Correlations with intelligence tests were low, though there was a slight tendency for the more intelligent man to be more radical and positive but less suggestible than the average.

*Accepted for publication by L. L. Thurstone of the Editorial Board and received in the Editorial Office, August 22, 1930.

To find out whether or not college freshmen really have definite religious convictions, a study was made by Dudycha (2) at Ripon College. During registration and before the students had attended a single college lecture, a series of 25 religious propositions was submitted to all freshmen. These propositions were stated in an affirmative manner and the students were asked to mark them according to the certainty of their belief or disbelief. About 25 minutes later, during which time they had been occupied with other tests, these students were given a second list containing the same propositions but expressed in a manner differently from the way the first series were stated. A correlation of .931 was obtained, from which the writer concluded that college freshmen are not misled by the phraseology of statements but respond according to their beliefs. Considering the group as a whole, not a single proposition was disbelieved more than it was believed.

In a study conducted by Betts (1) at Northwestern University to determine religious attitudes, students were asked to indicate whether they found the following factors in school life helpful or harmful: class instruction, campus religious organizations, fraternity or sorority relations, dormitory associations, campus religious opinion or attitude, personal friendships, campus public lectures, student paper. His findings indicated that whether a course was helpful or not depended principally upon the religious background and point of view the student brought to the course. Thirteen per cent judged that one or more courses had injured them religiously; 35% testified to aids from the different courses.

According to Lund (6), one believes most vividly what appeals to one's interests most strongly, although the thing believed may take a low place in the scale of certainty of belief. Poffenberger (7), in discussing conditions of belief in advertising, says, "We tend to believe what arouses our desires, our fears, and our emotions generally."

In an effort to elicit student attitudes on all important phases of campus and curricula situations, and to find what particular attitudes were apt to be found together in an individual student, Katz (5) submitted an anonymous questionnaire to students of Syracuse University. One important finding in his study was that attitudes on morality and on universality of cribbing seemed to go together with cribbing so decidedly that they might prove of value in gaining insight into the amount of cribbing behavior.

Following these various findings of students' opinions, a list of

25 statements on economic, civil, legal, religious, and social issues was devised. These 25 statements were submitted to 258 male college students at North Carolina State College. These students were in psychology classes and were distributed by classes as follows: freshman, 40; sophomores, 51; juniors, 105; and seniors, 62. This was an unselected group representing the college at large, since the classes in psychology from which these students came are elective. The students were given the following written instructions:

Below are given 25 statements. These statements are to be graded by you according to your beliefs of their being true or false. A grade of +2 indicates that you are strongly convinced that the statement is true, +1 a fair degree of certainty, 0 entire uncertainty, -1 that it is more wrong than right, and -2 that it is entirely false. You are to note the grade given by underlining the grade.

The list of the statements used is here presented (see Table 1). An examination of this list will reveal the fact that few if any of the statements can be answered with absolute assurance on the basis of verified facts now at hand. Furthermore, one will notice statements representing different fields of thought. These statements are not in general dependent upon technical knowledge and skill.

TABLE 1
STATEMENTS SUBMITTED TO STUDENTS

-2	-1	0	1	2	(1).	The United States has not yet reached its zenith.
-2	-1	0	1	2	(2).	Man was originally formed by a special creative act of God in a short time.
-2	-1	0	1	2	(3).	People who are unhappily mated should be allowed a divorce even when there is no adultery.
-2	-1	0	1	2	(4).	The plea of insanity should never free one from severe punishment when he has committed a horrible crime.
-2	-1	0	1	2	(5).	Prosperity, in general, is more pronounced during the administration of the Republican party.
-2	-1	0	1	2	(6).	Wars usually benefit only a few politicians and certain business interests.
-2	-1	0	1	2	(7).	Morality is on a higher plane today (in the U. S.) than it has ever been.
-2	-1	0	1	2	(8).	Christ was actually resurrected from the dead.
-2	-1	0	1	2	(9).	The present mode of dress is better for the moral welfare of society than was the mode 50 years ago.
-2	-1	0	1	2	(10).	Lynching is sometimes justifiable for horrible crimes, when the law is lax and non-expedient.

TABLE 1 (*continued*)

STATEMENTS SUBMITTED TO STUDENTS

-2	-1	0	1	2	(11).	The man who persists in exceeding the speed limit on the public highways should be jailed like other criminals.
-2	-1	0	1	2	(12).	Our government is controlled, in the main, by great financial interests.
-2	-1	0	1	2	(13).	The American worker's condition is better today than it was twenty-five years ago.
-2	-1	0	1	2	(14).	The Bible was divinely dictated to Moses and other apostles by God. These men wrote the words that were dictated to them.
-2	-1	0	1	2	(15).	It is harmless for young people who are not engaged to continually go night riding unchaperoned.
-2	-1	0	1	2	(16).	Movies and dancing are as apt to be morally good for anyone as morally bad.
-2	-1	0	1	2	(17).	Drinking coffee, when it is known to be injurious to the health, might be well considered a criminal act.
-2	-1	0	1	2	(18).	Poor men cannot get justice in the courts as they are ordinarily managed today.
-2	-1	0	1	2	(19).	Where religion is antagonistic to modern science religion should be allowed precedence.
-2	-1	0	1	2	(20).	"Peace on earth and good will to all men" seems near at hand at the present time.
-2	-1	0	1	2	(21).	The present status of marriage is in many respects much better than it has ever been.
-2	-1	0	1	2	(22).	A student in high school or college caught cheating in an examination should be suspended from school.
-2	-1	0	1	2	(23).	The expert mechanic should get about the same salary as the successful business executive.
-2	-1	0	1	2	(24).	The young of today are no more immoral than the young two or three generations ago were.
-2	-1	0	1	2	(25).	The soul of man has the power to control his actions independently of the physical circumstances in the world now and in his past.

Do not sign your name.

Year in College	College	Sex
-----------------------	---------------	-----------

Table 2 presents the total percentage of belief in each of the statements as expressed by the classes combined. The statement receiving the highest average positive score, indicating belief by the largest number of students, was No. 13, which refers to the American worker's condition today as compared with that of 25 years ago. Only 5% of the students think his condition unimproved during the last quarter of a century, while 3.9% are entirely uncertain.

TABLE 2

PERCENTAGE OF BELIEF IN EACH STATEMENT AS EXPRESSED BY 258 COLLEGE STUDENTS FROM ALL CLASSES

Statement No.	+2	+1	Rating 0	-1	-2
1	65.5	19.4	3.9	5.8	5.4
2	24.8	14.0	25.2	12.0	24.0
3	60.5	24.4	3.9	6.6	4.7
4	29.5	21.3	5.8	21.3	22.1
5	12.8	12.0	29.8	9.7	35.7
6	41.9	22.1	7.8	13.2	15.1
7	24.0	27.9	18.6	16.3	13.2
8	58.1	11.6	21.7	4.3	4.3
9	41.9	22.5	8.5	13.6	13.6
10	40.3	20.2	5.0	13.2	21.3
11	34.5	24.4	7.0	19.8	14.3
12	33.3	32.2	7.4	17.4	9.7
13	72.5	18.6	3.9	2.3	2.7
14	20.5	14.3	30.2	13.2	21.7
15	36.0	15.9	8.9	15.5	23.6
16	62.8	17.4	7.0	7.8	5.0
17	12.8	20.2	11.2	19.0	36.8
18	39.9	31.4	4.7	15.5	8.5
19	9.3	15.9	37.2	15.5	22.1
20	9.3	17.1	16.7	23.3	33.7
21	34.1	23.3	16.3	16.3	10.1
22	32.6	20.5	14.7	15.1	17.1
23	22.5	20.5	15.9	22.9	18.2
24	53.1	17.1	10.9	7.8	11.2
25	17.8	12.0	23.6	13.6	32.9

The social questions also rank high in percentage of positive belief. They are statements 3, 16, and 24, dealing with divorce, movies and dancing, and the probable immorality of the young people of today as compared with that of the young of two or three generations ago. Evidently by indicating belief in these statements, students are merely expressing the opinion of the enlightened majority of today. Statements 19 and 20 were given positive ratings by only 9.3% of the students. It is perhaps only natural, in this age of science, that students should not endorse No. 19, which states that religion should be allowed precedence over science in case of antagonism.

As to average certainty of opinion, it appears to be little affected by college courses. Comparison of the freshman with the senior class on the ratings +2 and -2, which indicates positive belief in the truth or falsity of statements, shows a difference of only .1%. In the positive rating the freshmen were 1.5% higher than the seniors,

while in the negative rating the seniors were 1.6% above the freshmen. However, there was a great variation between these two classes as expressed in the rating 0, given to statements about which students were entirely uncertain. This rating shows the greatest progressive increase in percentage from the senior class to the freshman, the difference between these classes being 3.4%. There was a gradual decrease in uncertainty from the freshman, who were 2.3% above the average for the classes combined, through the sophomore and the junior classes, to the senior which was 1.1% below the average. This decrease in uncertainty may be due partly to greater maturity and increase in general experience, and partly to the effect of college training upon the more advanced students. The juniors and seniors were slightly more positive that certain statements as expressed were false, while the freshmen and sophomores were more inclined to accept the statements as true; from which it might be concluded that the more mature students were less influenced by phraseology, and less suggestible than the younger.

Table 3 gives the average deviations for the ten statements showing the highest variability. It is interesting to note the wide deviation from the average. On 21 of the statements there was an average deviation of more than 1, while on all except one of the remaining four it was close to 1. The thirteenth statement, which it has already been pointed out received the highest positive score, showed the least variability. Considering the many different attitudes present in regard to law enforcement, it is not surprising that half of the statements listed in Table 3 as receiving the highest average deviation would be concerned with legal issues. The one statement on which there was most variability is No. 15, a social one, which indicates that college students in this age of the automobile and night-riding are divided in their opinion as to the harmlessness of this pastime. The statements showing the least variability are 13, 1, 3, 8, and 19 respectively.

TABLE 3

AVERAGE DEVIATION OF THE TEN STATEMENTS HAVING THE HIGHEST DEVIATION

Statement No.	<i>A.D.</i>	Statement No.	<i>A.D.</i>
15	1.50	17	1.33
10	1.47	22	1.33
4	1.46	6	1.32
25	1.45	9	1.28
11	1.34	23	1.27

Beliefs once formed are not usually very easy to change. The side of the question first presented to us and the early influences brought to bear upon us are very effective in determining our beliefs, so much as to suggest the presence of a law of primacy in persuasion (6, p. 195). This thought is in harmony with the findings and implications set forth in the investigations conducted by Jones (4) and by Betts (1). A more careful examination of the ratings given the various beliefs by classes tends to point to the same general conclusion.

Table 4 presents a comparison of the ratings of the freshman and senior classes on the statements involving some phase of a general religious thought. The religious statement most generally accepted by the 258 students was No. 8, which reads: "Christ was actually resurrected from the dead." Approximately 70% of all the students accept this statement or at least are inclined to believe it. Eight and five-tenths per cent gave it a negative rating, while 21.7% were entirely uncertain. Greater uncertainty was shown in the case of statement No. 19 than any other statement, students evidently doubting and in fact being inclined to disbelieve the idea that religion should have precedence over science in case of antagonism. In connection with this, it is very interesting to note their belief that religion should not have precedence over science and yet they endorse religious statements which are not compatible with the more generally accepted modern biological concepts concerning the creation, life, and general behavior of man. The other four of the religious-like be-

TABLE 4
COMPARISON OF FRESHMAN AND SENIOR CLASSES ON RELIGIOUS STATEMENTS
(Percentage Basis)

Statement No.	Class	Ratings				
		+2	+1	0	-1	-2
2	Fresh.	40	13	25	05	17
	Senior	21	13	29	08	29
8	Fresh.	68	05	17	05	05
	Senior	42	15	26	08	09
14	Fresh.	32	13	32	08	15
	Senior	19	13	26	13	29
19	Fresh.	08	15	53	12	12
	Senior	08	16	32	15	29
25	Fresh.	25	13	30	10	22
	Senior	13	16	19	10	42

liefs were more generally accepted by the freshmen than by the seniors. In the case of statement No. 25, which refers to the soul of man having power to control his actions, there is a decided shift from the positive to the negative ratings as the student passes from the freshman to the senior year.

Table 5 gives a still further comparison of the two classes on socio-economic questions. An analysis of this table will show that the seniors expressed decidedly less uncertainty on the majority of these issues than did the freshmen. On three statements (Nos. 7, 13, and 15) the two classes are comparatively close together in percentage of uncertainty. Only on proposition 17, which is surely a debatable question, were the seniors noticeably more uncertain than the freshmen, who gave this a negative rating. On statements 3, 9, and 21

TABLE 5
COMPARISON OF FRESHMAN AND SENIOR CLASSES ON SOCIO-ECONOMIC
STATEMENTS
(Percentage Basis)

Statement No.	Class	+2	+1	Ratings 0	-1	-2
3	Fresh.	45.0	32.5	5.0	12.5	5.0
	Senior	71.0	22.6	1.6	1.6	3.2
5	Fresh.	12.5	20.0	40.0	2.5	25.0
	Senior	9.7	16.1	30.6	11.3	32.3
7	Fresh.	32.5	22.5	12.5	12.5	20.0
	Senior	25.8	33.9	12.9	17.7	9.7
9	Fresh.	25.0	25.0	12.5	17.5	20.0
	Senior	50.0	22.6	4.8	14.5	8.1
12	Fresh.	20.0	30.0	10.0	27.5	12.5
	Senior	38.7	30.6	1.6	24.2	4.8
13	Fresh.	77.5	12.5	5.0	5.0	0.0
	Senior	64.5	25.8	6.5	3.2	0.0
15	Fresh.	37.5	27.5	7.5	12.5	15.0
	Senior	30.6	17.7	6.5	17.7	27.4
16	Fresh.	62.5	10.0	12.5	10.0	5.0
	Senior	62.9	24.2	6.5	4.8	1.6
17	Fresh.	7.5	15.0	7.5	27.5	42.5
	Senior	11.3	21.0	12.9	14.5	40.3
21	Fresh.	25.0	25.0	20.0	15.0	15.0
	Senior	32.3	25.8	17.7	17.7	6.5
23	Fresh.	27.5	15.0	30.0	2.5	25.0
	Senior	19.4	21.0	6.5	30.6	22.6
24	Fresh.	52.5	17.5	12.5	5.0	12.5
	Senior	46.8	17.7	8.1	9.7	17.7

the attitude of the advanced class was more liberal than that of the first year students in college. Nos. 3 and 21 refer to divorce and marriage, while statement No. 9 refers to the present mode of dress as compared with that of fifty years ago.

There seems to be a slight inconsistency, however, on the part of the freshman class for these ratings, or at least for part of them, just as was the case for the religious ratings. This inconsistency running through the freshman class ratings is probably due to lack of information and a narrower range of experience. It is but natural to expect uncertainty or a disbelief in the statement referring to the pronounced prosperity of the country under Republican administration, when this study was made in a state with a normal Democratic majority in political alignment. The seniors are rather well convinced that our government is controlled by great financial interests, since approximately 70% of them gave statement No. 12 a positive rating and only 1.6% expressed entire uncertainty. The freshmen are greatly divided in opinion on this particular issue.

A further analysis of the various beliefs portraying optimism, such as beliefs Nos. 1, 7, 9, 13, 16, 21, and 24, will show a decided amount of optimism present among the college students.

As a general summary, one might conclude that on the average there is no significant difference between the different classes in certainty of beliefs, but that a more detailed study by classes for the different beliefs revealed class difference in religious and socio-economic issues. It should be observed that North Carolina State College is a technical institution drawing her students primarily from the state. Furthermore, the growth of certainty or uncertainty of specific beliefs cannot be attributed wholly to the effects of college work on the students; but when one considers the types of opinions that seem to change during the college years, one is inclined to conclude that such changes are due to a more definite type of instruction than would ordinarily be gotten from incidental experience. The tendency on the part of the upper classmen to be more cautious and less suggestible, along with the greater amount of liberalism present in their attitudes on religious and social issues, would seem to point out some effect of training.

REFERENCES

1. BETTS, G. H. Religious attitudes and activities of university students, a report. *Relig. Educ.*, 1928, **23**, 917-919.

2. DUDYCHA, G. J. The religious beliefs of college freshmen. *School & Soc.*, 1930, **31**, 206-208.
3. GUTHRIE, E. R. Measuring student opinion of teachers. *School & Soc.*, 1927, **25**, 175-176.
4. JONES, E. S. Opinions of college students. *J. Appl. Psychol.*, 1926, **4**, 427-436.
5. KATZ, D. Student opinion at Syracuse. *Person. J.*, 1928, **7**, 103-110.
6. LUND, F. H. The psychology of belief. *J. Abn. & Soc. Psychol.*, 1925, **20**, 174-196.
7. POFFENBERGER, A. T. The conditions of belief in advertising. *J. Appl. Psychol.*, 1923, **7**, 1-9.
8. THURSTONE, L. L. Attitudes can be measured. *Amer. J. Sociol.*, 1928, **33**, 529-554.

North Carolina State College
Raleigh, North Carolina

UNE ÉTUDE DES OPINIONS DES ÉTUDIANTS UNIVERSITAIRES

(Résumé)

Après un résumé de quelques-unes des études les plus récentes et scientifiques des opinions des étudiants les investigateurs nous ont donné les résultats d'une étude des opinions des étudiants à North Carolina State College. On a soumis une liste de vingt-cinq déclarations sur des problèmes économiques, civils, légaux, religieux et sociaux à 258 étudiants mâles universitaires. On n'a pas choisi les étudiants et on les a distribués par années de la manière suivante: première année, 40; deuxième année, 51; troisième année, 105; et quatrième année, 62. On a évalué ces déclarations sur une échelle variant de +2 à -2; une évaluation de +2 indiquant une certitude presque complète de la vérité d'une déclaration, +1 un assez bon degré de certitude, 0 une incertitude complète, -1 plus fausse que correcte, et -2 que la déclaration est absolument fausse.

La certitude moyenne des opinions semble peu influencée par les cours suivis à l'université; bien que pour les déclarations spécifiques des différences constantes ont été exprimées par les groupes des étudiants des années première et quatrième. La déclaration recevant l'évaluation positive la plus élevée a été celle-ci: "La condition de l'ouvrier américain est meilleure aujourd'hui qu'il y a vingt-cinq ans." Les étudiants ont moins varié dans leur évaluation de celle-ci que dans celle de n'importe quelle autre déclaration. Les déclarations de nature religieuse orthodoxe en type ont été évaluées beaucoup plus libéralement par les étudiants de quatrième année que par ceux de première année. Pour les questions sociales un "air général" d'optimisme s'est manifesté. Pour les déclarations socio-économiques les étudiants de quatrième année ont exprimé décidément moins d'incertitude que ceux de première année. Plus de contradictions se sont montrées dans les croyances des étudiants de première année que dans celles des étudiants de quatrième année.

GARRISON ET MANN

EINE UNTERSUCHUNG ÜBER DIE MEINUNGEN VON STUDENTEN

(Referat)

Nach Revidierung der neuesten und wissenschaftlichsten Untersuchungen über die Studentenmeinungen bieten die Forscher die Resultate einer Untersuchung über die Meinungen von Studenten an der North Carolina State College dar. Eine Liste bestehend aus 25 Behauptungen über ökonomische, bürgerliche, gesetzliche, religiöse, und soziale Fragen wurden 258 männlichen Studenten an der Universität angeboten. Diese Studenten waren nicht ausgewählt. Sie waren auf die verschiedenen Klassen folgenderweise verteilt: Neulinge, 40; Studenten im zweiten Jahr, 51; Studenten im dritten Jahr, 105; Studenten im vierten Jahr, 62. Diese Behauptungen wurden nach einem Massstab geordnet, welcher sich von +2 bis -2 erstreckte. Die Zahl +2 deutete an, dass die Behauptung als fast bestimmt wahr galt; +1 bedeutete mittelmässige Bestimmtheit, 0 völlige Ungewissheit, -1 dass die Behauptung als mehr falsch wie wahr, und -2 dass sie als völlig falsch galt.

Die mittlere Gewissheit der Meinung scheint durch die Universitätsvorlesungen wenig beeinflusst zu werden, obwohl sich in Bezug auf bestimmte Behauptungen zuverlässige Unterschiede zwischen den Neulingen und den Studenten der vierten Klasse erwiesen. Die Behauptung welche den höchsten positiven Rang erhielt lautete: "Der Zustand des Amerikanischen Arbeiters ist heutzutage besser, als er vor fünfundzwanzig Jahren war." Die Studenten variierten in ihrer Ordnung dieser Behauptung weniger, als bei irgend einer anderen. Die Behauptungen religiöser Natur wurden von den Studenten der vierten Klasse viel freisinniger rangiert als von den Neulingen. Bei sozialen Fragen zeigte sich ein allgemeiner Ton des Optimismus. In Bezug auf soziale Fragen erwiesen die Studenten der vierten Klasse entschieden weniger Unbestimmtheit als die Neulinge. Die Meinungen der Neulinge zeigten mehr Widersprüche als die der Studenten der 4ten Klasse.

GARRISON UND MANN

RACIAL PREFERENCES OF A THOUSAND AMERICAN UNIVERSITY STUDENTS*

From the Department of Psychology, University of Nebraska

J. P. GUILFORD

This study of racial preferences was begun with the primary purpose of testing the method of paired comparisons as applied to the measurement of racial attitudes. As a secondary problem, it was conducted to see how students in widely scattered parts of the country would rate fifteen of the races and nationalities¹ which have made up our American population and which still ask for admission by way of immigration. With the cooperation of a number of psychologists, ratings were secured from about eleven hundred students in seven different universities and colleges as widely separated as New York University, the University of Florida, and the University of Washington.² The results were found to be so reliable that a number of additional problems have been illuminated by statistical treatment of the data which we have secured. If the experimental setting has been somewhat defective for the solution of all these problems, this report may at least point the way to a scientific attack and even indicate in what direction the answers probably lie.

Previous to the winter of 1928, when this study was begun, there had appeared only a few preliminary attempts to measure racial attitudes as a form of "social distance," a concept that has attained considerable vogue in sociology (1, 2, 3, 4). Only a roughly defined rating-scale had been used up to that time. The problem of racial likes and dislikes, prejudices and antipathies seemed important

*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office June 19, 1930.

¹The term "race" will be used in a non-technical sense in this paper for the sake of brevity. The groups with which we are dealing are more properly called "cultural groups," or even nationalities, instead of races.

²Acknowledgments are made to the following individuals who generously assisted in securing data: Professor Beulah Morrison, University of Kansas; Professors A. R. Gilliland, Franklin Fearing, and S. N. Stevens, Northwestern University; G. B. Vetter of New York University; C. O. Weber of Wells College; Osborne Williams and H. A. Swenson of the University of Florida; R. H. Gundlach of the University of Washington.

enough to warrant the application of that most precise psychometric method, the method of paired comparisons. Since this study was begun, a similar one has been published by Thurstone which fully justifies the method as a tool of measurement in this field (12). The latter report leaves untouched most of our problems, however, so the present study was brought to its final conclusion.

PROCEDURE

Our method of securing the data is similar to that of Thurstone except that he used the names of 21 races as stimuli, whereas we have used only 15. The 15 are as follows: Chinese, English, French, German, Greek, Hindu, Italian, Japanese, Jew, Mexican, Negro, Russian, Spanish, Swedish, Turk. The terms "Hindu" and "Jew" were used because they have a more definite meaning to the average person than "Indian" and "Hebrew." These names were arranged in pairs, as in the method of paired comparisons, 105 pairs in all, and presented in mimeographed form to the students in classes in psychology. The following instructions were printed at the top of the page which contained the 105 pairs, and were read to the subjects before they were permitted to begin:

Instructions: Below is a list of different nationalities arranged in pairs. Underline in each pair the one whom you would rather admit to your country as a fellow citizen. *Be sure to underline one in each pair.*

At the bottom of the sheet was a space for recording the father's and mother's races or nationalities.

The smallest number of subjects at any school was 62 at Wells College and the largest was 211 at the University of Kansas. It was found that even the smallest group gave exceedingly highly reliable ratings of the 15 races. We believe, because of the high reliability, that we have an accurate measurement of the student opinion at each university. The coefficients of reliability will be found in Table 2. The exact number of subjects at each university will be found in Table 1.

After the data were secured, the method of computing scale values differed from that of Thurstone. The method is shorter and simpler. It has been fully described by the writer in an earlier article (7) and later justified empirically in a second report (8). It must be regarded only as an empirical method which works.

Briefly stated, it consists of the following steps. In any group of data from the method of paired comparisons one determines the total number of choices given to any particular race, plus a small correction. This total is divided by the total number of times that the race appeared for choice in all its combinations. This gives the "mean proportion" for that race. The difference between this mean proportion and .50 is computed. Then, by reference to a table of the normal probability function, one finds the linear distance on the abscissa that corresponds to that difference. This distance may be either in terms of the standard deviation as the unit or in terms of the probable error. We have used the latter, because it is equivalent to the old and familiar j.n.d. of psychophysics.

The scale values for each race as measured at each university will be found in Table 1. The final column gives the unweighted averages of the scale values from the six universities exclusive of New York University. The latter gives results so divergent from the others in some respects that they have been treated separately in most of the discussion which follows. The negative values indicate that some races are rated below the average of the group. A negative sign does not necessarily mean a dislike for any race, therefore, since we do not know just where on the scale the true indifference point lies. We could, of course, shift the zero point arbitrarily so as to make all the values positive, but the writer believes that they are more significant left the way they are. Needless to say, a race that is preferred less than half of the time will have a negative scale value.

RELIABILITY OF THE RESULTS

It does not require very close inspection of Table 1 to see that there is considerable harmony in racial attitudes among the different universities, with the exception of N. Y. U. First, let us ask regarding the agreement within each university. How well do the students at the same school agree with each other? In order to determine this, the results from each university were divided into two equal parts, selecting papers at random, and two sets of scale values were computed. A product-moment coefficient of correlation was then found between these two sets of scale values. By using the Spearman-Brown prophecy formula, the reliability of the scale values for the combined data at each university was determined. These coefficients range from .967 for Wells College, to .9995 for the University of Kansas, with all but one above .995. It appears that

TABLE 1
SCALE VALUES OF THE FIFTEEN RACES AT THE SEVEN UNIVERSITIES

N	Fla. 198	Kans. 211	Nebr. 110	N. Y. U. 100	N. U. 208	Wash. 207	Wells 62	All ex. N. Y. U. 996
Chin.	-1.207	-1.282	-979	-876	-1.083	-1.174	-795	-1.085
Eng.	2.346	2.334	2.168	1.589	2.253	2.288	2.417	2.302
Fr.	1.310	1.371	1.246	1.032	1.288	1.344	1.328	1.314
Ger.	1.310	1.365	1.537	1.182	1.354	1.246	1.384	1.366
Gk.	-.202	-.469	-.518	-.493	-.302	-.394	-.386	-.377
Hindu	-1.188	-.674	-.955	-1.048	-.745	-1.294	-.750	-.932
Ital.	.224	.134	.148	.096	.246	.223	.147	.165
Jap.	-.649	-.733	-.462	-.540	-.688	-.898	-.603	-.676
Jew	-.217	-.346	.045	1.310	-.197	-.138	-.379	-.205
Mex.	-.628	-.616	-.912	-.740	-.750	-.639	-.605	-.693
Negro	-.757	-1.078	-.892	-.938	-1.017	-.750	-1.165	-.944
Rus.	-.064	.251	.187	.491	.111	.287	.139	.151
Span.	.600	.554	.318	.059	.437	.367	.321	.433
Swed.	1.106	1.238	1.305	.348	1.132	1.209	1.206	1.200
Turk.	-1.348	-1.326	-1.440	-1.218	-1.424	-1.322	-1.501	-1.395
Range	3.694	3.660	3.608	2.807	3.677	3.610	3.918	3.697
σ <i>dist.</i>	1.055	1.082	1.050	.910	1.036	1.072	1.066	1.060

TABLE 2
INTERCORRELATION AND SELF-CORRELATION OF THE SCALE VALUES

	Fla.	Kans.	Nebr.	N. Y. U.	N. U.	Wash.	Wells
Fla.	.998	.982	.978	.853	.989	.991	.978
Kans.	.982	.9995	.979	.847	.995	.982	.989
Nebr.	.978	.979	.997	.894	.987	.979	.975
N. Y. U.	.853	.847	.894	.995	.866	.877	.843
N. U.	.989	.995	.987	.866	.998	.982	.991
Wash.	.991	.982	.979	.877	.982	.998	.976
Wells	.978	.989	.975	.843	.991	.976	.967
r_{ag}	.990	.990	.994	.871	.998	.993	.986
$(r_{ag})^2$.980	.980	.988	.759	.996	.986	.972

100 representative subjects would have been amply sufficient from each university to secure an adequate sampling of student opinion on the question of racial preferences.

COMMUNITY OF RACIAL OPINION

The intercorrelations are nearly as high as the self-correlations, with the exception of those involving N. Y. U. They range from .843 to .894 for N. Y. U., but they are between .975 and .991 for all the other universities. Surely, this means a very great unanimity of opinion among all the thousand students excepting those of N. Y. U.

Let us assume that the high positive coefficients do represent a common element or a complex of common determiners for opinion in all the universities, including N. Y. U. Let us assume, further, that the lack of perfect correlation indicates local determiners of opinion, peculiar to the one university alone. In the case of N. Y. U. it is not difficult to surmise what this local factor is. Seventy-one per cent of the subjects at this university have Jewish parentage. There is also a cosmopolitan factor present in this group that is absent in most of the others. It is to be admitted that the students at each university are not necessarily local products. While the bulk of the students at each *state* university in the list come from within the boundaries of that state, and while outsiders who come to those universities do tend, to some extent, to take on local coloring in their attitudes, there are too many opinion-forming forces of nation-wide scope, like the radio, the movie and the press, to permit purely local attitudes to develop. And probably the racial attitudes of students are determined to a much greater extent by these indirect factors

than they are by first-hand contacts with any of the races or nationalities in these experiments.

Since the intercorrelations *are* somewhat lower than the self-correlations, it argues for some local determiners of racial attitude. We may assume, for the sake of argument, that the determiners of these attitudes thus fall into two groups, like Spearman's "*g*" and "*s*" factors (11). There are no good apparent reasons why there should be any "group" factors aside from "*g*" to account for the high intercorrelations. Such "group" factors might appear, for example, among urban and rural groups of students, and within different social or economic classes. The test for "group" factors will be made later. For the present, we have assumed that only "*g*" and "*s*" determiners exist, and we have found the correlations of each set of scale values with this "*g*" (11, appendix p. xvi). The "*g*," under this assumption, represents the consensus of opinion of all seven groups. The correlations of the opinion of each group with the consensus of opinion are given at the bottom of Table 2 under the heading of "*r_{ag}*." They are all above .986, with the exception of the coefficient of .871 for N. Y. U. Assuming, further, that the amount of overlapping of the two measures—the opinion of the group and the consensus of opinion—is measured by r^2 , we can show to what extent the attitudes of any one group are determined by that common factor (6, p. 294). The percentage of "*g*" represented in the scale values for each university is given in the bottom row of Table 2. The percentages vary from .759 in the case of N. Y. U. to .996 for Northwestern. If our original assumption of a "*g*" factor and "*s*" factors only is correct, this means that in most universities the common determiner, or determiners, account for 98% of the expressed opinion on races.

Perhaps the application of Spearman's technique in this problem needs justification. So far as the writer knows, this is the first attempt to use his method in this way. We have assumed, in doing so, that each university is analogous to a mental test and that each race is like an individual who takes that test. It will be interesting to see what the results are when the tetrad difference criterion is applied to the table of intercorrelations. If more universities were included, or if a number of more homogeneous groups of subjects were used, interesting group factors in racial attitudes should appear. The same technique should apply to the analysis of all kinds of attitudes, political, religious, or social, to help determine the degree of una-

nimity upon any public question and to detect the lines of cleavage of public opinion.

While we are comparing the seven groups of subjects, let us examine Table 1 more carefully to note any outstanding generalities or discrepancies that may have significance. It is to be noticed that all groups, even N. Y. U., are decidedly pro-English, and the average distance from the English to the next highest nationality, the German, is almost one j.n.d. The lowest race in every list is the Turk. Undoubtedly the Turk owes his position on the scale to our lack of first-hand contacts with him and to the overwhelmingly unfavorable impressions gained through indirect sources. This is borne out by the results from one student who had lived in Turkey for a number of years who preferred the Turk to all others except the English. All races are rated consistently positive or negative, except the Russian in the case of Florida, and the Jew in the case of Nebraska and N. Y. U.

There is some solace for those who possess the "Nordic complex" in the fact that most Nordic races stand high in the scale. Not one falls below the average in the final values, but of the southern Europeans, the Italians and the Spanish are always on the positive side. All Asiatics are consistently negative, with the Japanese higher in the scale than the Hindus or the Chinese.

MEASURES OF TOLERANCE AND HETEROGENEITY

Thurstone has made the interesting suggestion (12, p. 416) that we can measure the "international mindedness" or tolerance of any group of subjects by the total range of scale values which it gives. This is based upon the assumption that if the subjects had few real preferences, their judgments would often be matters of chance, and hence proportions near .50 and small separations between scale values would result. All races would cluster about the zero point. Thurstone's hypothesis would be true if tolerance were the only factor which produces a small range of proportions in a group. The writer doubts that it is.

We find the total ranges in scale values for the seven groups of subjects in this study at the bottom of Table 1. They are rather uniform, but they vary from 2.807 to 3.918 j.n.d.'s. If Thurstone's interpretation is correct, the results mean that N. Y. U. shows the greatest degree of tolerance and the Wells College girls the least. Now the total range of measures is the roughest indicator of dis-

person. It is limited to two values, the two extremes. We have therefore computed the standard deviation of the distribution of the scale values about the mean as a more accurate measure of dispersion. These values are given in the very last line in Table 1. Again, N. Y. U. has the smallest dispersion, but now Kansas has the greatest, not Wells College. Several other changes in order of "tolerance" occur when we use the standard deviation instead of the range.

But does a small variability in the scale values from their mean indicate tolerance? One might the more readily acquiesce to this interpretation after noting the results. Surely New Yorkers are noted for tolerance and Kansans have been less well-known for that trait. But the writer believes that another and more important factor is responsible. It is the heterogeneity of the group. Imagine a group of subjects so homogeneous that every pair of races is nearly always judged the same by every subject. The proportions would be nearly 0 or 1.00 in every case. Imagine, on the other hand, a group so heterogeneous that one-half of the subjects rate the races in the opposite direction to the other half. Proportions near .50 would result. The group at N. Y. U. is known to be very heterogeneous, as Table 4 will show. Its scale values have the smallest standard deviation. The group at Kansas, which has the highest standard deviation, was probably more homogeneous than any of the other universities because it draws so large a proportion of its students from Kansas alone. From the writer's impression, the group at Kansas was very tolerant. Some subjects at that university thought it absurd or impossible to express any racial preferences at all. And would not a group with racial antipathies new from Europe, as one would find in New York, be less tolerant than a midland population several generations removed from Europe?

Wells College is also high in the list for homogeneity due to the fact that the subjects were all of the same sex. Surely these girls, many of whom have traveled abroad, should have an "international" attitude. Northwestern comes next to N. Y. U. for heterogeneity—or tolerance. A good case can be made for heterogeneity here. Northwestern is not a state university, it draws both rural and urban students, and it has considerable racial mixture among the parents (see Table 4). It might be, of course, that the heterogeneity of the group could be interpreted as "tolerance," that is, the heterogeneous group will welcome into itself more readily a racial variety of new-

comers, but certainly the individuals in it need not be tolerant and in fact might be just the reverse.

It is of interest to compare these results with certain laws of social distance as formulated by Poole and Poole (10). Only those laws which are pertinent to our results will be considered. One of them should be mentioned at this point. Law 5 states that "social distance increases with increasing segregation." No one, I think, will dispute the relationship. But it may work both ways. A race that is generally disliked is forced to segregate itself, whereupon lack of contact and understanding further increases the gap. Table 1 shows what we would expect, that the most segregated groups in our country fall toward the lower end of the scale. Especially is this true of the Negro and the Chinese and to some extent of the Jew and the Greek.

VARIABILITY OF RACIAL PREFERENCES

It was noted in tabulating the data that some races tended to occupy about the same place in the series of 15 from one subject to another, whereas other races varied over the whole range. For example, the French may have consistently received 12, or nearly 12 choices, out of a possible 14, whereas the Japanese received all the way from 1 to 10 choices from time to time. This can mean only one thing: that the group, as a group, is quite certain of the position of some races on the scale and quite uncertain of others. It is of interest to know of which races they are more certain, or upon which they agree best, and of which races they are least certain. Is there more agreement upon those races near the ends of the scale and less upon those near the middle? Is there more agreement upon those that are liked than upon those that are disliked? We would expect, *a priori*, positive answers to both these questions. Disagreement of the members of the group upon any race would produce proportions in the direction of .50. Close agreement would pull the proportion away from the mean in either the positive or negative direction. Extreme racial prejudices, favorable or unfavorable, for those races that are generally disliked would cause a relatively wider scatter for those races than for those that stand high in the scale. This should be so because the races high in the list have a generally recognized merit which, under the instructions of the experiment, guarantee a universal high rating even by those who may dislike them. In other words, the preferred races may be judged more objectively and the "unpreferred" ones more subjectively, i.e., more on the basis of feeling.

The customary way of measuring the variability of each race on the scale would be to find the probable error of its scale value. Unfortunately, so far as the writer knows no method has been devised for getting a probable error for such measurements. A method, which might not bear strict mathematical scrutiny, has been improvised for the occasion. We do know, for any group of subjects, just how many preferences each subject gave to each race. The number of preferences may vary from none to 14, since each race is judged 14 times. We can therefore secure a frequency distribution of the number of subjects who chose a given race 14, 13, 12, etc., times. By inspection, these distribution curves seem to be nearly always of the unimodal normal type. Those distribution curves for races near the extremes of the scale are skewed, because no more than 14 choices nor no less than 0 choices could be given. It was decided, therefore, to find the quartile deviation (Q) rather than the probable error. These values of Q are presented for each race at each university in Table 3. N. Y. U. differs so much from the other six universities again, that a mean of only the latter six was secured and these means are found in the eighth column of Table 3, labeled Q_1 .

This table is worthy of minute inspection. There are a number of significant differences in the Q 's of the different universities in spite of the general agreement. One can only surmise the reasons in each case, and, with the information we have, one guess is about as good as another. Here are a few outstanding single cases: The excep-

TABLE 3
Q-VALUES FOR EACH RACE

	Fla.	Kans.	Nebr.	N. Y. U.	N. U.	Wash.	Wells	Q_1	Q_2	Q_1/Q_2
Chin.	1.46	1.49	1.91	2.13	1.58	1.39	1.80	1.60	1.41	1.13
Eng.	.35	.34	.60	.91	.36	.43	.35	.40	.55	.74
Fr.	.94	.82	.89	1.15	.81	.87	.89	.87	1.24	.70
Ger.	.98	1.05	1.00	1.36	.98	1.04	.93	1.00	1.20	.83
Gk.	1.60	1.51	1.46	1.36	1.37	1.41	1.51	1.48	1.78	.83
Hindu	1.29	1.88	1.70	1.74	1.63	1.59	1.67	1.63	1.50	1.09
Ital.	1.32	1.33	1.38	1.67	1.53	1.36	1.62	1.42	1.82	.78
Jap.	2.01	1.90	2.06	2.30	2.03	1.72	2.15	1.98	1.65	1.20
Jew	2.48	2.16	1.82	1.40	2.81	2.22	2.53	2.34	1.81	1.29
Mex.	1.52	1.72	2.00	1.78	1.76	1.59	1.95	1.76	1.63	1.08
Negro	2.48	2.16	2.61	1.83	2.22	2.09	1.86	2.24	1.49	1.43
Rus.	1.91	1.43	1.67	1.44	1.61	1.30	1.56	1.58	1.82	.87
Span.	1.09	1.17	1.57	1.83	1.30	1.25	1.36	1.29	1.76	.73
Swed.	1.05	.98	.93	1.62	1.09	.99	.80	.97	1.32	.73
Turk	1.20	1.34	1.46	1.64	1.41	1.14	1.28	1.30	1.18	1.11

tionally high Q 's for the Chinese and the Italian at Wells College; for the Hindu at Kansas; the Russian at Florida; the exceptionally low Q for the Japanese, the Russian, and the Turk at Washington. Students at Washington seem to have more definite opinions of the value of Asiatics, as we might expect. Another of Poole's laws of social distance is *a propos* here (10, p. 366). Law 1 states, "Continued and intimate contact reveal the true social distance." This law would seem to say that repeated contact with a race makes us more certain of its position on the scale. We have no measurement of the degree of contact with these few races at the above universities, so we can only suggest that the law applies to these cases.

As for general tendencies, Table 3 is not disappointing. Two general tendencies were predicted, and they were found. The one was that when compared with the scale values the Q 's should tend to increase near the mean of the scale and to decrease toward either end, and the other was that there would be a negative correlation between the Q 's and the scale values.

The first of these two general tendencies should be so because of the following mathematical considerations. Whatever method will finally be devised for finding the probable errors of scale values which are derived by the method of paired comparisons will have to take account of the fact that the proportions near .50 have the largest probable error and those near 0 and 1.00 have the smallest. The general formula, $P.E._p = .6745\sqrt{pq/N}$, in which p is the experimental proportion and q is equal to $1-p$ and N is the number of cases. The writer has used this formula as a basis for computing some theoretical Q 's from the proportions of choices received by each race. These probable errors were numerically too small to compare with the observed Q 's, but they were adjusted to the proper size by multiplying them by a common factor of 4.6. These theoretical Q 's will be found in Table 3, in the next to the last column. In Figure 1 we have both the Q_1 and Q_2 values plotted against the scale values. A line is drawn through the theoretical (Q_2) values. If continued, this line would be in the form of the normal distribution curve, with its mode at the scale value of zero where the observed proportions are .50. The actual Q values are scattered about this curve, some above and some below.

The actual Q values now have more meaning. Except for three races, Jew, Negro, and Jap, they do follow the expected normal distribution fairly well. This is not entirely due to the fact that those

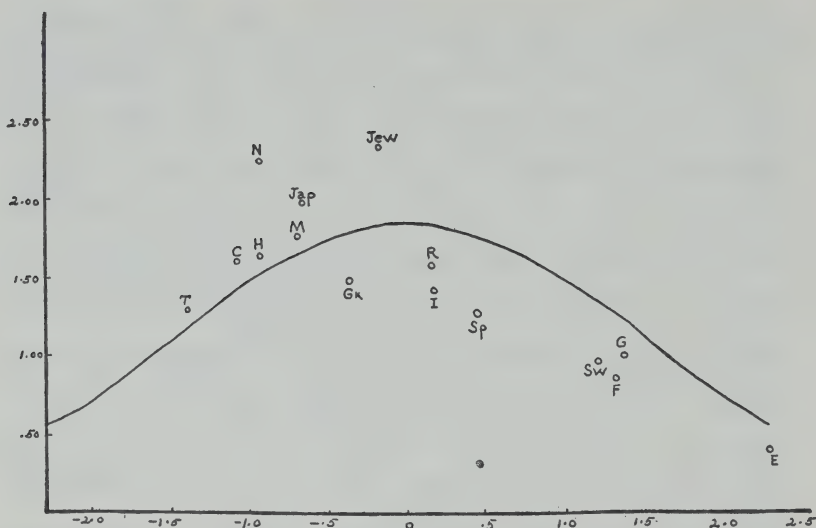


FIGURE 1

Showing the relationship between the scale values and their absolute variabilities as measured by the median deviations of the number of choices given to each race. The curve represents the "best-fitting" normal distribution of the expected probable errors of the scale values.

(Abcissa - - - Scale values)

(Ordinate - - - Median deviations)

racers near the end of the scale cannot be chosen more than 14 times nor less than zero, for we have used Q instead of the probable error. The normal distribution demonstrates that the formula for the probable error of a proportion will have to be considered in finding a probable error of the scale values.

Figure 1 shows to some extent that our second prediction was not far wrong. There is a definite negative correlation between the Q 's and the scale values. Only one race with a minus scale value is below its theoretical value. This relationship will be seen more clearly if we reduce the theoretical Q 's to a single level. We have, accordingly, the ratios between Q_1 and Q_2 in the last column of Table 3 and also a graphic picture of these ratios as compared with the scale values, in Figure 2. Imagine the normal curve of Figure 1 flattened out into the straight line which is drawn at ordinate 1.00 in Figure 2. We now have a truer picture of the relation of these Q 's to their scale values than we had before. In Figure 1 the Jew had the highest absolute Q ; he is the hardest to place by the thousand

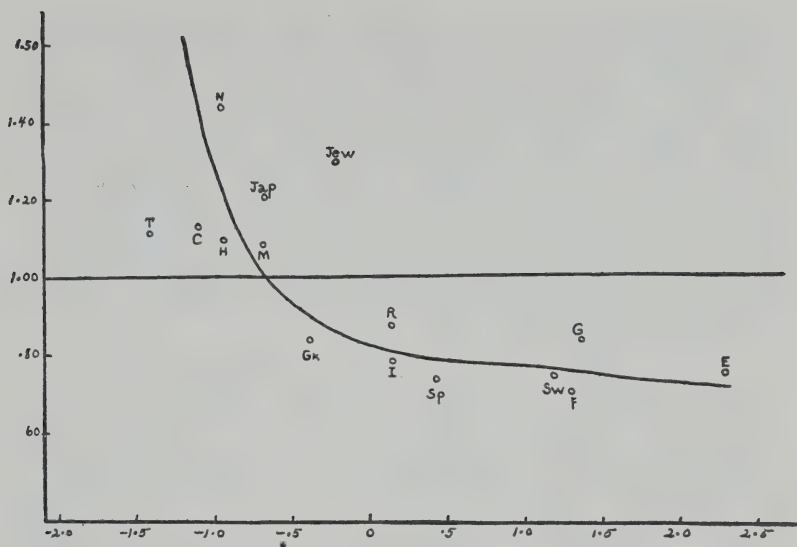


FIGURE 2

Showing the relationship between the scale values and their relative variabilities as measured by the ratio of the observed median deviations to the theoretical probable errors.

(Abscissa - - - Scale values)

(Ordinate - - - Ratios of variability)

students. But in Figure 2 we see that the Negro is the most uncertain quantity, relative to his scale value. In Figure 1 the Englishman seems by far the easiest to place on the scale, but in Figure 2 it will be seen that, relative to their scale values, the Spanish, French, and Swedish are equally easy to allocate.

All in all, this seems to be a fairly good indicator of race prejudice, if prejudice is defined as judgment over-determined by feeling. We do have a wide variety of feeling about the Negro, the Jew, and the Jap, but probably for different reasons in each case. In the case of the Jew it may be the wide variation in the members of the race itself. We meet such extreme samples, from the standpoint of economic and social status and of personal qualities. This may account for our mingled feelings toward that race. With the Negro, another explanation may be necessary, and so, with the Jap. This involves the whole problem of race prejudice which is too large a subject to be handled here.

The writer does wish to point out, however, the very singular relationship between the relative Q 's and the scale values of the races. A curve has been drawn through the points by inspection. The regression is clearly non-linear, but of what type and of what significance the writer does not venture to say. Repeated experiments under better analyzed conditions are needed to test for some quantitative law which surely holds for these data. Hollingworth has found similar results in judgments of likes and dislikes (9, p. 103). A group of subjects was more consistent with judgments of stimuli near the top of the scale than near the bottom. Whether or not this is a universal tendency, I do not know.

DETERMINERS OF RACIAL PREFERENCES

Nothing has been said so far in this report, except in a very incidental way, concerning the determiners of the scale values. It is impossible, with the limited data we have, to make any complete or sure analysis of the determining factors. But I do wish to indicate some promising methods for so doing and to suggest some possible results. In the discussion which follows liberal use has been made of methods of correlation. This procedure may be of doubtful value with so few cases, but the results seem to tell a reasonable story. The probable errors of the coefficients are given in each case to remind the reader of their relatively low reliability.

The first factor to suggest itself is the race of the parentage of the students who did the choosing. We asked each subject to state his father's race and also his mother's. Because of the kind of results which we have obtained from this request, we must be cautious. It is likely that some students gave *American* as the nationality of their parents if they were the first ancestors they had who were born in this country, whereas others gave German or English even when the family had lived on American soil for several generations. Again, a very few took the instructions literally, interpreting "race" in the technical sense, and simply responded with "white," which means nothing for our purposes. Many subjects also gave races not included in our list of 15. These additional racial descents may have their influences upon the preferences of some students, but we cannot know what effect nor how much without careful analysis. An interesting study would be to group the subjects according to racial ancestry and to find the scale values for each group. We have not done this as yet.

Sixty-five per cent of the 996 subjects gave ancestry from one or more of the 15 races involved, however, so that by neglecting the minority of 35 per cent we still have a fair idea of the racial make-up of the combined group. We did not eliminate the 35%, but will assume that their preferences are proportional to those of the majority with whom they live and associate.

Table 4 gives the number of subjects claiming full parentage from each of the 15 races at each university. By full parentage is meant the equivalent of four grandparents of a particular race. If, for example, a student gave his father's race as English, and his mother's race as German and French, we gave the English two points and the French and German each one point. The total number of points for any race was divided by four to reduce the number to the number of subjects who have the equivalent of full parentage in that particular race.

A further study of this factor should include a very careful check upon the race of the parents, so that one group of subjects may be compared with another on that basis. It is apparent that such comparisons may be significant when we note in Table 4 the relatively large number of Jews and Russians represented at N. Y. U. and, corresponding to this, the relatively high scale values for the two races at N. Y. U. The Russian parentage represented at N. Y. U. happens to be composed mainly of Russian-Jews which were counted twice, once as Russians and once as Jews.

Other differences between the universities in Table 4 are so small that such comparisons cannot be continued profitably. In the last column of the table will be found the percentage of the 996 subjects who give parentage in each of the 15 races. These may be compared with the average scale values with significant results. The product-moment coefficient of correlation between parentage and the scale values for 12 of the races, excluding the Jews, Mexicans, and Negroes, is .814, with a probable error of .113. These three races were omitted from the computation because they have no fixed immigration quotas and later we wish to bring in comparisons with the immigration quotas. The relationship between the parentage of the subjects and the scale values for all 15 races is illustrated in Figure 3. A curve has been drawn through the points by inspection. The regression is clearly non-linear, but a correlation ratio cannot be found since there are too few cases. The coefficient of correlation is therefore smaller than it should be, both because of the non-linear

regression and because of the fact that the scale values are determined in part by the 35% whose parentage is not among the 15 races concerned.

An examination of Figure 3 is much more illuminating than a coefficient of correlation can be with so few cases as this. The decided positive acceleration of the curve should be apparent. It is reminiscent of certain logarithmic curves encountered in psychophysics. Something like Fechner's law may apply here. But when we plot the scale values against the logarithms of the parentage of each race, the regression is still non-linear. If we assume the function to be the square root, and if we correlate the scale values with the square roots of the parentage, r is $.888 \pm .090$. If we use the fourth root of the parentage, r is but slightly higher, but it is $.940 \pm .067$. We may have a logarithmic function in these data, but it is not the simple function of the Fechnerian law. It is probably near the fourth root. The uncertainty of the data makes it useless to attempt to find the exact function. We shall meet similar fourth root laws again, however, in comparing the scale values with the immigration quotas.

Assuming that the fourth root law holds, we have the problem of explaining deviations from it. These are patent in Figure 3. The Germans have too low a scale value and the French too high in proportion to the German and French parentage in the group. This may as well be blamed upon the recent war, since so many other things are also charged against it. The relatively high value given to the Swedish race is inexplicable, unless it is the fact that it is a Nordic race. The Italian and Russian groups are also higher than we would expect from the law which we have assumed. These two races are very poorly represented among the parentage of our students, but they make up a very large proportion of our present foreign-born population according to the 1920 census (see Table 5). It will be shown later that there is some slight positive correlation between the 1920 foreign-born population and the scale values apart from the parentage of these students.

The case of the Spanish cannot be accounted for so easily. The number of foreign-born Spanish in the country has always been small. The relatively high value placed upon the Spanish must be either a tribute to desirable qualities recognized by the students—possibly the study of the Spanish language has had something to do with it—or else the Spanish owe their high rating to the fact that

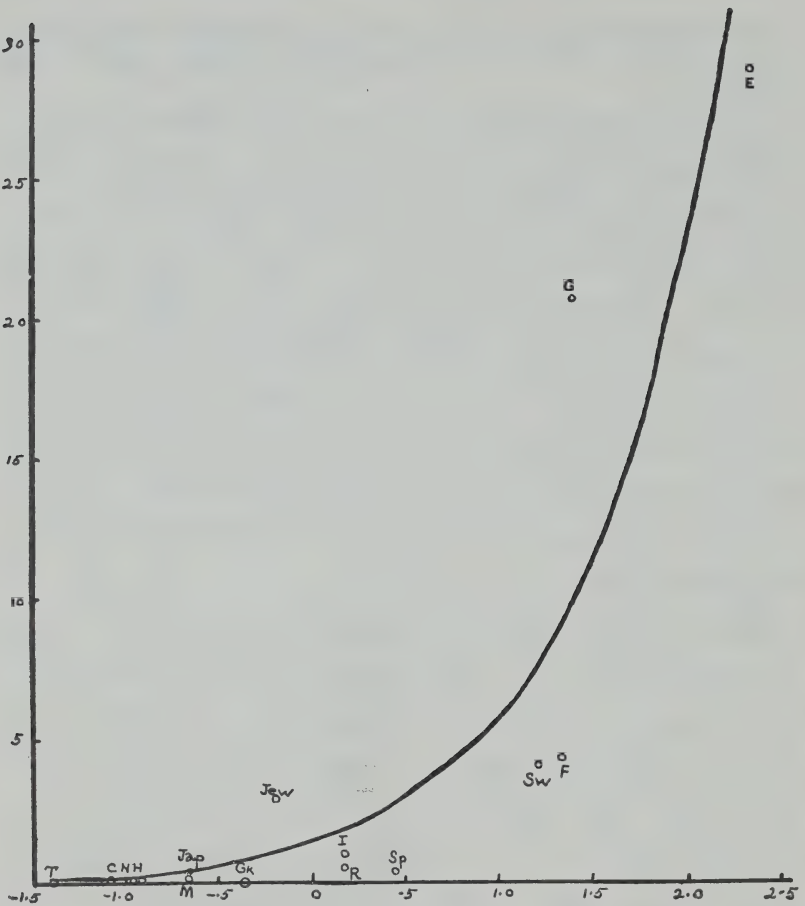


FIGURE 3

Showing the relationship between the scale values and the racial composition of the group of subjects.

(Abscissa - - - Scale values)

(Ordinate - - - Parentage, in percentages)

they are compared with less desirable races. The latter explanation could also apply to the Italian, Russian, and Swedish as well, but we have also suggested some more positive reasons for their high ratings.

It would be pertinent to ask whether the parentage of this group of 996 students is representative of the racial origins of Americans

of this student generation. In order to answer this question we have recourse only to the United States census reports. We will assume that the parents of these students were counted in the census reports of 1890, 1900, and 1910. We can find the number of foreign-born of each race in the country at those times and secure an average of each as an index of that race. One fact makes this comparison of doubtful value. Many of the subjects gave foreign ancestry which dates far back of 1890 and the foreign-born population has been shifting since 1890 to some degree. The correlation between the parentage of our group and the foreign-born population from 1890 to 1910 is to the extent of $.875 \pm .094$; very significant, but not perfect. An examination of the data shows this lack of perfect correlation is due to the relatively small number of Italian, Russian, and German ancestry among our students as compared with the relatively high proportion of French and English ancestry represented there. It is very likely that this French and English ancestry dates back beyond 1890, and the Italian, Russian, and perhaps the German does not. Aside from this, it tells us which races are contributing relatively more or less to our American college population.

Table 5 presents the indices of the foreign-born populations from 1890 to 1910 for the 12 races in question. It also shows the same for the 1920 census. In both cases, columns three and four, the numbers represent thousands of foreign-born individuals of each race. The immigration quotas for 1929 and 1930 are also given in the table. We shall use them presently.

TABLE 5

Race	Scale values	Parentage (per cent)	Foreign-born (thousands)		Quotas	
			1890-1910	1920	1929	1930
Chin.	-1.087	.00	82	44	100	100
Eng.	2.302	29.20	1636	1700	34007	65700
Fr.	1.314	4.25	111	153	4000	3100
Ger.	1.366	20.40	2700	1686	51227	26000
Gk.	-.378	.00	63	170	100	300
Hindu	.934	.00	3	5	100	100
Ital.	.165	1.00	670	1610	3845	5800
Jap.	-.676	.30	32	81	100	100
Rus.	.151	.68	854	1400	2248	2800
Span.	.433	.45	12	50	131	250
Swed.	1.200	4.20	575	625	9561	3300
Turk	-1.394	.00	31	16	100	220

In spite of the fact that the races, as represented by the foreign-born population from 1890 to 1910, have not contributed proportionately to the student population, should they not, by their presence, by the cultural milieu with which they have surrounded these subjects as children, influence their preferences? We may bring in at this point Poole's third law of social distance, which states that "personal distance increases as spatial distribution increases" (10). The converse of this law may not be true, but we might test it with these data. Stated thus, the law would say that the closer we live with a race the less the social distance; the better we like them. There are good reasons for believing that this does not always hold. If there is competition or incompatibility between races living close together, the social distance may be increased.

At any rate, the correlation between the scale values and the foreign-born population for each race from 1890 to 1910 is only $.656 \pm .147$. And if we find the partial coefficient, holding parentage constant, r is even negative, $-.198 \pm .191$. This would seem to mean that racial likes and dislikes are but little influenced by the social milieu of one's childhood alone. Their roots run deeper than that. This is an unexpected result and one which should be more carefully tested. The relationship between the 1890-1910 populations and the scale values is non-linear, however, so this partial coefficient may not mean anything at all. It would be small, in any case, because of the relatively high correlation between the parentage and the foreign-born population of 1890-1910.

Since we found the scale values proportional to the *fourth root* of the parentage, we might also expect them to be more closely proportional to the fourth root of the 1890-1910 foreign-born populations. This is true. The correlation in this case amounts to $.775 \pm .123$. It is probably illegitimate to secure a partial coefficient in this instance, but if we do, it is more negative than before: r is $-.281 \pm .187$.

Will the present associates, or rather the present fellow citizens, of these students have any effect upon their preferences? We do not have the 1930 census reports available, but we will assume that the foreign-born population has not changed materially since 1920, and use those figures. In this case, the correlation with the scale value is $.632 \pm .152$. The partial coefficient, with parentage held constant, is a positive $.139 \pm .193$. This partial coefficient could apparently be almost anything. The regression is non-linear again, and the correla-

tion between the scale values and the fourth roots of the foreign-born populations of 1920 is $.827 \pm .109$. The partial coefficient, with parentage held constant, is a positive $.668 \pm .144$. This is quite a different story, if true. It would mean, if the coefficient were valid and significant, that the present racial environment of the subjects does carry some weight, independent of the ancestry of their parents.

RELATION OF PREFERENCES TO IMMIGRATION QUOTAS

During the last ten years Congress has wrestled with the problem of the restriction of immigration. Two problems have faced the country: to stem the tide of immigration from Europe which set in following the close of the War, and to decide which races should be given preference. It is reasonable to suggest that, in a representative form of government, the preferences of the people themselves should have something to say in determining the source of their newest adopted countrymen. A rating device such as the method of paired comparisons would be an excellent plan for the purpose. In view of the unanimity of opinion found among our 996 students, it would not be a prohibitive task to sound the public opinion of the country on this question.

Assuming that the 996 students give us a fair sampling of American opinion as to which races should be preferred in framing the immigration quotas, let us see how well the present quotas agree with their racial evaluations.

The 1929 quotas, as given in Table 5, are based upon the proportions of the foreign-born populations in the 1890 census, except that no race has a quota of less than 100. The 1930 quotas are based upon the "national origins" plan (5, Chap. 8). The writer has been unable to find the exact basis for these quotas, but gathers that a complex system of computation has been used to find the proportion of each race which has gone into the making of the present American population. Since we found a high correlation between the scale values and the foreign-born populations from 1890-1910 (r equals $.775$ with the fourth root of the latter), we would expect a correspondingly high correlation between the scale values and the 1929 quotas which are based upon the 1890 census. In this case r is $.690 \pm .141$. The regression is clearly non-linear (see Figure 4), so we have correlated the scale values with the square roots of the quotas. The coefficient is then $.812 \pm .114$. But the plot is still non-linear. Using the fourth root of the quota, the correlation amounts to $.820 \pm .111$.

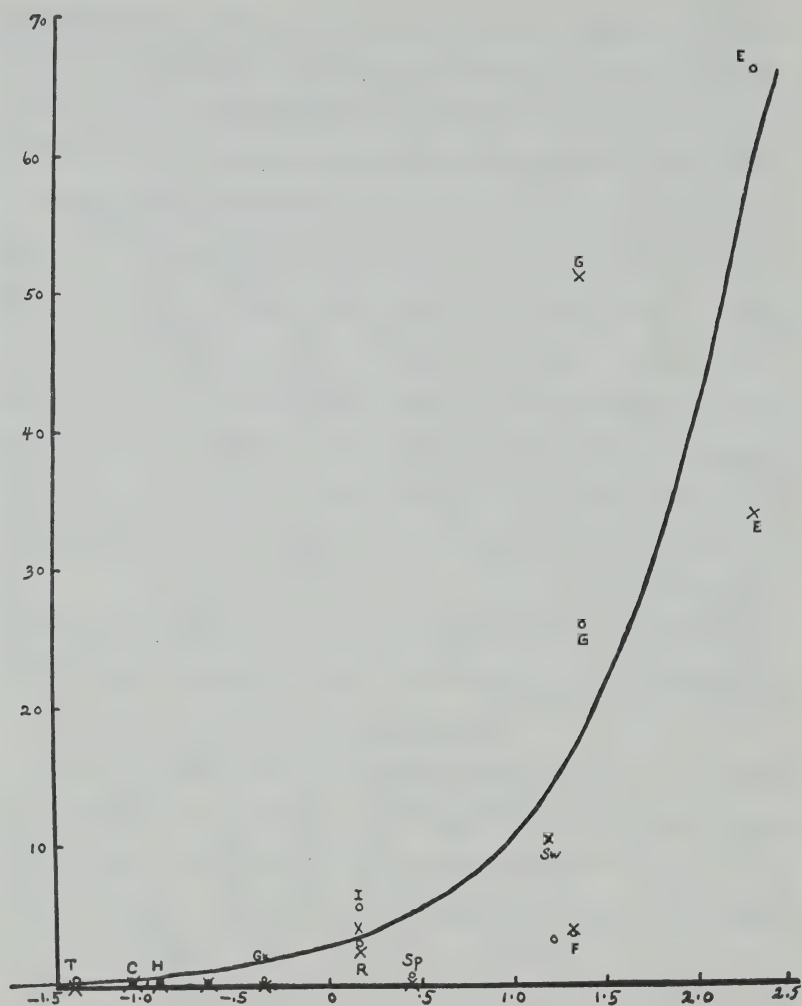


FIGURE 4

Showing the relationship between the scale values and the immigration quotas for 1929 and 1930. The 1929 quotas are represented by crosses; the 1930 quotas by circles.

(Abcissa - - - - Scale values)

(Ordinate - - - - Immigration quotas)

The 1930 quotas correlate somewhat better with the scale values than do the 1929 quotas. The coefficient is $.708 \pm .138$, if we use the quotas themselves; $.810 \pm .114$, if we use the square roots of the quotas; and $.918 \pm .078$ if we use the fourth roots. In both the 1929 and 1930 quotas the "fourth root law" seems to hold. Some races approach the curve more closely in 1929 and others in 1930. The German quota is a little high in both cases, but we must remember that the scale value of the Germans is low in comparison with the parentage of the students who rated them, and still more in comparison with the foreign-born populations from 1890 to 1920. The English quota was decidedly too low for 1929, but perhaps too high in 1930. The Swedish, French, and Spanish quotas could be higher than they are, especially in 1930. The Russian quota is exactly right, but the Italian, which was nearly correct in 1929, was given an unwarranted raise in 1930.

The judgments which have just been passed upon the immigration quotas are based upon the assumption that the fourth root law does hold between popular preferences and the quotas, and also that the preferences of the thousand students are representative of the country at large. Both these assumptions may be questioned, but not the essential procedure for finding the preferences and applying the results. It may be that another law should hold, but we found that this same fourth root law held between the scale values and the parentage of the judges. The subjects are clearly not representative of the 1920 population. They contain too few Russians and Italians, especially. Whatever weight popular preferences should have in framing the immigration quotas, however, the method of paired comparisons will furnish a very reliable index of those preferences.

SUMMARY AND CONCLUSIONS

Of the following, conclusions 8 and 9 can be given only tentative support:

1. Fifteen races, nationalities, or "cultural groups" were rated by the method of paired comparisons upon their desirability as American citizens. The judges were approximately 1000 university students.

2. A very high reliability was found for the scale values of the races at each university.

3. Very high intercorrelations were obtained between the scale values at the different universities. The community of opinion was about 98% for six of the universities and 76% for the seventh.

4. A method of measuring the "heterogeneity" of opinion of a group was suggested and applied to the seven groups in this study.

5. Opinions of races near the extremes of the range were less variable than those near the center.

6. Opinions were the more variable, the lower the scale value.

7. A method of measuring the degree of prejudice for each race was suggested. Prejudice was defined as judgment which is over-determined by personal feeling.

8. The strongest determinant of racial evaluation seemed to be the racial ancestry of the subjects. The scale values were proportional to the fourth root of the racial ancestry, or to some function near the fourth root.

9. There seemed to be very little relationship between the racial environment of the group and their preferences, if racial ancestry is partialled out.

10. There was a significant correlation between the scale values and the 1929 and 1930 immigration quotas, especially with the fourth roots of those quotas.

REFERENCES

1. BOGARDUS, E. S. Measuring social distance. *J. Appl. Sociol.*, 1925, **9**, 299-308.
2. ———. Social distance and its origins. *J. Appl. Sociol.*, 1925, **9**, 216-226.
3. ———. Sex differences in racial attitudes. *J. Soc. & Soc. Res.*, 1927, **12**, 279-285.
4. ———. Race friendliness and social distance. *J. Appl. Sociol.*, 1927, **11**, 272-287.
5. GARIS, R. L. Immigration restriction. New York: Macmillan, 1927.
6. GARRETT, H. E. Statistics in psychology and education. New York: Longmans, Green, 1926. Pp. 330.
7. GUILFORD, J. P. The method of paired comparisons as a psychometric method. *Psychol. Rev.*, 1928, **35**, 494-506.
8. ———. Some empirical tests of the method of paired comparisons. *J. Gen. Psychol.*, 1931, **5**, 64-77.
9. HOLLINGWORTH, H. L. Experimental studies of judgment. *Arch. Psychol.*, 1913, No. 29. Pp. 119.
10. POOLE, W. C., & POOLE, H. K. Laws of social distance. *J. Appl. Sociol.*, 1927, **11**, 365-369.
11. SPEARMAN, C. Abilities of man. New York: Macmillan, 1927. Pp. xxiii+415.
12. THURSTONE, L. L. An experimental study of nationality preferences. *J. Gen. Psychol.*, 1928, **1**, 405-423.

University of Nebraska
Lincoln, Nebraska

PRÉFÉRENCES DE RACE DE MILLE ÉTUDIANTS UNIVERSITAIRES AMÉRICAINS

(Résumé)

On a évalué quinze races, nationalités ou "groupes de culture semblable" par la méthode des comparaisons mises en paires au point de vue de leur capacité de devenir des citoyens américains désirables. Les juges ont été 1100 étudiants universitaires choisis parmi sept universités de différentes parties des Etats-Unis. On a computed des valeurs d'échelle pour chaque race à chaque université. Les corrélations de soi de valeurs d'échelle indiquent une constance extrêmement élevée des évaluations. Des corrélations entre elles-mêmes entre les valeurs d'échelle aux différentes universités indiquent une grande unanimité des opinions sur les 15 races. On a estimé que l'accord des opinions a été environ 98 pour cent pour six universités et 76 pour cent pour la septième.

Les opinions des races près des extrêmes de la rangée ont été moins variables que celles près du milieu. En général, moins la race est désirable, plus les opinions ont varié. On a suggéré des méthodes de mesurer l'hétérogénéité d'un groupe de juges et aussi de mesurer les préjugés de race pour n'importe quelle race donnée.

Le plus fort facteur déterminant des préférences de race a été la race des ancêtres des juges. Les valeurs d'échelle ont été proportionnelles à la quatrième racine de la race des ancêtres. On n'a trouvé nulle ou presque nulle relation entre le milieu de race des juges et leurs préférences, si la race des ancêtres reste constante. Il y a une corrélation positive très significative entre les valeurs d'échelle et les cotes de l'immigration en 1929 et en 1930. Les valeurs d'échelle ont été proportionnelles à la quatrième racine des cotes de l'immigration.

GUILFORD

DIE RASSENBEVORZUGUNGEN VON TAUSEND STUDENTEN AN AMERIKANISCHEN UNIVERSITÄTEN

(Referat)

Fünfzehn Rassen, Nationalitäten, oder Kulturgruppen wurden nach der Methode der gepaarten Vergleichen (paired comparisons) in Bezug auf den Grad ihrer Erwünschtheit (desirability) als Amerikanische Bürger rangmässig geordnet. Die Richter waren 1100 Studenten gewählt aus sieben über die Vereinigten Staaten zerstreut liegenden Universitäten. Es wurden für jede Rasse an jeder Universität Rangwerte (scale values) ermittelt. Die Intrakorrelation (self-correlation) dieser Rangwerte deuten sehr hohe Zuverlässigkeit der Abschätzungen an. Die sehr hohe Inter-korrelationen zwischen den Rangwerten an den verschiedenen Universitäten erweisen starke Uebereinstimmung der Meinungen über die 15 Rassen. Der Uebereinstimmungsgrad der Meinungen (community of opinion) liess sich für sechs Universitäten als ungefähr 98% und für die siebente al ungefähr 76% abschätzen.

Meinungen über jene Rassen die nah an den Enden der Rangordnung standen zeigten im Allgemeinen um so kleinere Abweichungen, desto weniger wünschenswert die Rasse. Es wurden Methoden vorgeschlagen, zur Messung der Ungleichartigkeit einer Richtergruppe und auch zur Messung des Vorurteils gegen irgend eine besondere Rasse.

Als das wirkungsvollste Element bei der Bestimmung des Rassenvorurteils zeigte sich die Abstammung der Richter. Die rangmässigen Abschätzungen ("scale values") erwiesen sich als proportionell zur vierten Wurzel (the fourth root) der Rassenabstammung. Es zeigte sich wenig oder kein Verhältnis zwischen der Rassenumgebung der Richter und ihren Vorurteilen wenn Rassenabstammung konstant gehalten wurde. Man fand eine sehr bedeutungsvolle positive Korrelation zwischen den rangmässigen Abschätzungen und die Einwanderungs Quoten für 1929 und 1930. Die rangmässigen Abschätzungen erwiesen sich als proportionell zur vierten Wurzel der Einwanderungsquoten (immigration quotas).

GUILFORD

RACIAL COMPARISONS OF ABILITY IN IMMEDIATE RECALL OF LOGICAL AND NONSENSE MATERIAL*¹

From the Psychological Clinic, University of Hawaii

C. M. LOUTTIT

Studies on racial differences have been directed largely to investigations with individual or group intelligence tests. Specific topics such as reaction-time, sensory acuity, memory, attention, and the like have been relatively neglected. As it seemed desirable that investigation be started on some of these topics, this study of memory, which plays such a large part in performance on most intelligence tests as well as in school, was undertaken. The problem was to determine whether or not differences in immediate recall of both logical and nonsense materials exist between certain available racial groups.

I. METHODS

A. Material. The four tests used in the present study—marble statue, letter square, auditory memory span (digits), and visual memory span (digits)—were all suggested by Whipple's *Manual* (5), which source was followed also in scoring and treatment of data.

1. *Marble Statue.* This test is a simple story containing, according to Whipple's division, 67 logically related ideas. It was read slowly and clearly to each group of subjects. Immediately after the reading was completed the subjects began to write the story. In all cases they were advised that it was not necessary to reproduce *verbatim*, but to do their best to recall ideas. In scoring these papers generous allowance was made for grammatical errors, misspelling, etc., and credit was given if there was any indication of having the idea.

2. *Letter Square.* This test of memory for serial impressions contains elements, particularly spatial relations, that make it a somewhat more complex test than mere memory span, and yet it lacks the meaning element of the Marble Statue Test. The material

*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office, April 25, 1930.

¹This investigation was supported by the Rockefeller Fund for Studies of Race of the University of Hawaii. The work has been done under the auspices of the Psychological Clinic of the University.

consisted of three rows of four consonants which were exhibited visually, and immediately after their removal the subjects attempted to record both letters and positions correctly.

Table 1 shows the six combinations used. Each block was painted with velvet black show-card color on vellum finish wedding bristol cards. Each card, 16 by 20 inches, had one block centered on it. The letters were two inches square and were separated from their neighbors by two inches. This size was easily seen in all parts of the largest room used at any time. Card (S) was used as a sample in explaining the test, but the subjects were not required to report on it. The cards (I-V) were shown in order for 25 seconds each, and at least one minute allowed for making the record. In scoring, 3 was allowed for each letter in the correct position, 2 for each one only one place removed either horizontally or vertically, and 1 for each letter two places removed. Omissions, wrong letters, or letters more than three places removed were counted zero. The individual's score was the average of the scores for the five cards.

TABLE 1
LETTER SQUARE COMBINATIONS

(S)	LBNJ HSDR CXFK	(I)	MTDX VLGN SZBR	(II)	RNBQ ZLSV PDMH
(III)	ZBPF MGWC SXHT	(IV)	FCKP JQBL NHXR	(V)	SXTC KNFR BJLW

3. *Auditory Span.* The digits shown in the left-hand column of Table 2 were read aloud to a group of subjects at the rate of one digit per second. Starting with the group of five digits, the series were read in order to nine digits (ten in the case of the university students). Immediately following the reading of each group the subjects wrote them down.

4. *Visual Span.* The procedure was the same as for the Auditory Span Test except that cards containing the digits were exhibited for the same number of seconds as the card had numbers. These digits are given in the right-hand column of Table 2. The cards were 4 inches high with 2-inch figures spaced two inches apart.

The scoring for the Auditory and Visual Span Tests was the same. The score for memory span was taken to be the number of digits correctly reproduced at least once immediately preceding the first

TABLE 2
DIGITS USED IN MEMORY SPAN TESTS

Auditory span	Visual span
41863	13925
57296	57296
736142	241637
619538	374259
2417358	9537142
7392148	6415937
81539274	47293615
54263718	72594136
816537429	475296318
473692581	819753462
4052731968*	2046173958*
9527306418	8637192504

*Ten digits were given only to the university students.

complete failure. Thus if the subject got one of the pair of seven digits correct but failed both of the eights his score was seven.

B. Subjects. The subjects were all of the 12-year-old boys and girls of White, Japanese, Chinese and Hawaiian or Part-Hawaiian (hereafter designated Hawaiian) ancestry in seven public schools at Honolulu available on the day the test was given. In calculating the averages only 50 unselected cases were used for each of three of the sex-race groups. There were only 40 cases available for the Hawaiian groups.

All of the pupils to be used in a school were assembled in a convenient large room and the test given to them. These groups included 41, 65, 70, 76, 77, and 87 individuals at different schools, and each had representatives of at least two races. Each child was supplied with an examination "blue-book" for recording the responses.

In addition to school children, all of the tests were given to the writer's class in elementary psychology consisting of approximately 150 students. The tests were presented, in the same manner as described, during a regular class period. A few of the university students were discarded because they represented racial groups or mixtures other than those being studied.

II. RESULTS

A. Marble Statue Test. Pyle (3) has given the Marble Statue Test to school children in Canton, China. The test was translated into Chinese and given by a native teacher. His results for 12-year-

olds were: 22 boys scored 28.77 and 12 girls scored 37.00. Although he gives no results for adults, his data for 18-year-olds may be roughly compared to our university students. His scores are: for 20 boys 26.20, and for 8 girls 30.50. He compares these scores to the following white norms: 12-year boys 35.01, girls 37.79; 18-year boys 36.80, and girls 38.12. If these averages for whites are compared with those shown in Table 3, it will be seen that, except for 18-year-old girls, his averages are greater. Inasmuch as there is a judgment factor involved in marking this test, I am inclined to attribute a good share of the difference to that fact. On the other hand, for the Chinese groups my averages exceed his in all cases except the 12-year-old girls. This difference may be due to possible selection involved in the small number of his cases and possibly to the difference in language.

In Table 3 are shown the averages for the present investigation. It will be seen that the 12-year white males exceed all other groups; the Japanese, Chinese, and Hawaiian following in that order. The white females also exceed the others, but the Chinese now take second place, with the Japanese and Hawaiians following. For the university students the orders are: males—White, Chinese, Japanese; females—White, Hawaiian, Chinese, Japanese.

Because of the small number of cases it does not seem advisable to indicate the σ 's of the differences between the university groups. For the 12-year-olds these values for differences between the White and other groups have been calculated. These figures are: males—WJ 1.192, WC 1.404, WH 1.452; females—WJ 1.476, WC 1.482,

TABLE 3
AVERAGE SCORES ON THE MARBLE STATUE TEST

Ancestry	12-year-olds			University		
	Cases	Av.	σ	Cases	Av.	σ
Males						
White	50	32.96	6.55	6	35.33	3.45
Japanese	50	32.46	5.31	26	32.80	6.21
Chinese	50	30.94	7.47	10	33.40	7.31
Hawaiian	40	27.90	7.07	(*)		
Females						
White	50	33.64	7.27	25	43.40	7.39
Japanese	50	30.10	7.51	20	36.45	6.49
Chinese	50	31.88	7.53	24	37.91	6.33
Hawaiian	42	29.48	6.46	12	38.25	4.76

*Only 2 cases.

TABLE 4
AVERAGE SCORES ON LETTER SQUARE TEST

Ancestry	12-year-olds			University		
	Cases	Av.	σ	Cases	Av.	σ
Males						
White	50	22.24	5.90	6	25.56	4.87
Japanese	50	28.21	5.83	26	28.61	5.46
Chinese	50	26.12	6.68	14	28.82	4.28
Hawaiian	40	25.91	6.57	(*)		
Females						
White	50	21.10	5.30	23	26.64	5.02
Japanese	50	23.18	6.54	22	27.69	4.59
Chinese	50	22.56	6.91	22	27.25	6.43
Hawaiian	42	26.04	6.47	12	29.53	3.27

*Only 2 cases.

TABLE 5
AVERAGE AUDITORY SPAN

Ancestry	12-year-olds			University		
	Cases	Av.	σ	Cases	Av.	σ
Males						
White	50	6.98	1.39	6	7.00	1.52
Japanese	50	7.02	1.52	26	7.29	.99
Chinese	50	6.98	1.31	14	7.71	1.22
Hawaiian	40	7.05	1.45	(*)		
Females						
White	50	7.00	1.37	23	7.43	1.37
Japanese	50	6.80	1.34	22	7.18	1.02
Chinese	50	7.12	1.29	22	8.04	1.49
Hawaiian	42	7.21	1.21	12	8.00	.91

*Only 2 cases.

TABLE 6
AVERAGE VISUAL SPAN

Ancestry	12-year-olds			University		
	Cases	Av.	σ	Cases	Av.	σ
Males						
White	50	7.76	1.09	6	8.33	.94
Japanese	50	8.32	.97	26	8.42	1.24
Chinese	50	8.06	1.22	14	9.00	1.25
Hawaiian	40	7.65	1.33	(*)		
Females						
White	50	7.90	1.08	23	8.69	1.36
Japanese	50	8.14	1.72	22	8.63	1.29
Chinese	50	7.84	1.30	22	8.54	1.03
Hawaiian	42	7.64	1.41	12	9.25	1.09

*Only 2 cases.

WH 1.432, indicating respectively 6591, 9251, 9996, 9916, 8830, and 9981 chances in 10,000 that the differences are greater than 0. Only the difference between White and Hawaiian, which is 3.48 times its σ , can be considered significant. The differences between White and Japanese females and White and Hawaiian females, being between 2 and 3 times their σ 's, are possibly significant.

B. Letter Square Test. The norms published by Whipple for this test taken from unpublished data of Anderson and from Winch (6) differ from each other to a degree represented by more than one σ in the distribution discussed in this paper. Thus Winch's average of 31.1 and Anderson's of 15.7 are greater than 1σ plus and minus the averages for 12-year-olds given below. This disagreement makes it inadvisable to make comparisons.

The data from this research are shown in Table 4. Japanese males excel all other groups, with the Chinese, Hawaiian, and White following in that order. For the females, the Hawaiians take the lead, with the Japanese, Chinese, and Whites following. The same order holds for the university groups with the exception of both Japanese and Chinese males and females who are tied for second place.

Values of the $\sigma_{diff.}$ for the 12-year-old groups compared with the Whites are: males—WJ 1.720, WC 1.260, WH 1.332; females—WH 1.247, WJ 1.190, WC 1.231, indicating respectively 9999.5, 9599, 8810, 9995, 9989, and 9970 chances in 10,000 that the differences are greater than 0. The differences between White and Hawaiian males, White and Japanese females, and White and Chinese females, which are respectively 3.96, 3.37, and 3.08 times their σ 's, are significant. With the exception of the difference between White and Hawaiian males which is over twice its σ and therefore may be a real one, the other differences are too small to be considered of any significance. In all cases the differences that do exist favor the groups other than White.

C. Auditory Span. Smedley (4) gives the auditory memory span for digits as 6 for 12-, 6 for 18-, and 7 for 19-year-olds. Hao (2) has given this test to 600 Chinese school children in San Francisco with these results: males—12-year, 6.5, 18-year, 6.0; females—12-year, 5.9, 18-year, 6.2. The results of both these investigations are somewhat less than the averages secured in this study.

Table 5 shows the data of the present investigation. Twelve-year-old males of the four racial groups do not differ more than .07 in the extreme comparison. The university males show a little

greater variation than this but hardly enough to be significant. Likewise, the groups of females all vary less than one point. Calculation of the chances of the differences being greater than 0 shows that the greatest is only 7823 in 10,000. In five of the 12-year-groups the slight differences are in favor of the groups other than White, in one they are equal, and in the last the Whites excel (see Table 8).

D. Visual Span. For the visual memory span, also, both Smedley and Hao give data for Whites and Chinese respectively. The former found 7 digits for 12-, 7 for 18-, and 8 for 19-year-olds, while the Chinese 12-year males scored 8.2 and the 18-year-olds, 9.6; 12-year females scored 7.3, and the 18-year-olds, 8.0. These values compare more favorably with mine than do those for auditory span.

Data on the visual span are shown in Table 6. Like the differences for auditory span, none of these exceed one point. However, one of them may be significant, namely, that between White and Japanese males with 9967 chances in 10,000. This difference is 2.72 times its σ .

E. Comparison of the Four Tests. Each of the tests given had a finite range, that is, from 0 to a known upper limit. This makes it possible to express the performance as a percentage and to compare the four tests. Table 7 shows the percentage of the highest possible score for each race-sex-age-test group.

TABLE 7
PERCENTAGE SCORES

Ancestry	12-year-olds					University				
	Marble Statue	Letter Square	Auditory Span	Visual Span	Average	Marble Statue	Letter Square	Auditory Span	Visual Span	Average
Possible Score	67	36	9	9		67	36	10	10	
	%	%	%	%		%	%	%	%	
Males										
White	49	62	78	86	68.75	53	71	70	83	69.25
Japanese	48	76	78	92	73.50	49	80	73	84	71.50
Chinese	46	72	78	90	71.50	50	80	77	90	74.25
Hawaiian	42	72	78	85	69.25	*				
Females										
White	50	59	78	88	68.75	65	74	74	87	75.00
Japanese	45	64	76	90	68.75	54	77	72	86	72.25
Chinese	48	63	79	87	69.25	57	76	80	85	74.50
Hawaiian	44	72	80	85	70.25	57	82	80	93	78.00

*Only 2 cases.

These percentages indicate that the relative difficulty of the four tests may be represented by the descending order; Marble Statue, Letter Square, Auditory Span, and Visual Span, which is the order in which they were presented.

The averages of these percentages for each race-sex-age group give a rough measure of that group's ability in immediate recall. Such averages do not demonstrate the presence of any real racial difference in the ability. The university women in all groups did better than the 12-year-old girls, but this is not true of the men. Further, the university women did markedly better than the university men, although such a difference is not evident between the 12-year-old boys and girls.

III. DISCUSSION

The four tests used in this research have one element in common, namely, each requires a written report of material immediately recalled after one presentation. This common element appears in its purest form in the simple memory span tests, both auditory and visual. The first requires recall of relations only in a one-dimensional temporal series, and the latter in a one-dimensional spatial series.

Two factors, besides the common one, are important in the Marble Statue Test. These are (*a*) the use of language, English in the present case, and (*b*) the logical relations among the ideas in the passage and the meaning of the passage as a whole. The language factor probably plays some part in the efficiency of performance because the 12-year white groups, both boys and girls, who have a better acquaintance with good English, excel all other groups. Also the university students do considerably better than the 12-year-olds. This I attribute to their longer educational experience which materially assists both their language ability and their grasp of the relations of ideas.

In addition to the common factor, the Letter Square Test involves the apprehension of spatial relations in two dimensions. Here, again, the university students excel, not, however, to the degree evident in the Marble Statue Test, indicating that increased experience or maturity is of advantage. It is evident, therefore, that factors other than that of memory play fairly significant parts in performance on these tests.

Turning now to a consideration of the significance of the differences which do exist between the various 12-year-old groups, we find

TABLE 8
DISTRIBUTION OF $\frac{\text{diff.}}{\sigma_{\text{diff.}}}$ VALUES

diff. $\sigma_{\text{diff.}}$	Marble Statue	Letter Square	Auditory Span	Visual Span	Total
0			WCm		1
less than 0.99	WJm		WHf— WCf— WJf— WHm WJm—	WHf WCf WJf— WHm	10
1.00- 1.99	WCf WCm	WHf— WJf—		WCm—	5
2.00- 2.99	WJf WHf	WHm—		WJm—	4
3.00- 3.99	WHm	WHf— WCm— WJm—			4

m—male

f—female

pairs marked (—) are in favor of the group other than White.

that they are low. There are 24 comparisons with the white groups as shown in Table 8. In 66.6% of these the differences are less than twice the σ 's, indicating no significance. Four comparisons, or 16.6%, have the differences 2 to 2.99 times as large as the σ 's, suggesting that these may be real differences. The balance, 16.6%, have the difference three or more times the σ 's, or, accepting the usual rule of statistical procedure, are significant. Of these four differences that are significant, one occurs in the Marble Statue Test, three in the Letter Square and none in either the Auditory or Visual Memory Span Tests. This would suggest that factors other than memory are operating. Further evidence of this is seen when we consider the comparisons of the Auditory and Visual Span Tests. None of the auditory test comparisons are significant and only one of the visual may possibly be so. Because of the small number of cases, the $\sigma_{\text{diff.}}$ values for the university groups have not been calculated. However, examination of the data for these groups indicates that they closely parallel those for the 12-year-olds.

The percentage scores shown in Table 7 also indicate the lack of real differences between the race groups excepting only those whose

$\frac{diff.}{\sigma_{diff.}}$ values are greater than 3. The lack of racial differences is even more positively shown by the averages of the percentage scores, which I take to be adequate composite measures of performance in the four tests.

From the evidence here presented, it appears that there are no real differences between the racial groups studied, in fact the differences between sex groups and age groups are of greater significance. Those few differences that are large enough to be significant may be due largely to other factors than memory.

IV. SUMMARY

1. The Marble Statue, Letter Square, Auditory Digit Span and Visual Digit Span Tests were given to groups of 12-year-old school children and to university students, both male and female, representing White, Japanese, Chinese, and Hawaiian ancestry.

2. Comparison of the various racial groups with Whites shows small and insignificant differences. Only 10 of the 24 comparisons were in favor of the Whites.

3. Differences between sex and age groups are larger than those between race groups.

REFERENCES

1. ANDERSON, E. J. [Unpublished data quoted extensively by Whipple (5).]
2. HAO, Y. T. The memory span of 600 Chinese school children in San Francisco. *School & Soc.*, 1924, **20**, 507-510.
3. PYLE, W. H. A study of the mental and physical characteristics of the Chinese. *School & Soc.*, 1918, **8**, 264-269.
4. SMEDLEY, F. W. Child study in Chicago. Chap. 27 in the Report of the U. S. Commission of Education for 1902, Vol. 1. Pp. 1094-1138.
5. WHIPPLE, G. M. Manual of mental and physical tests. Part II. Complex processes. Baltimore, Md.: Warwick & York, 1915. Pp. v+336.
6. WINCH, W. H. Immediate memory in school children. *Brit. J. Psychol.*, 1904, **1**, 127-134; 1906, **2**, 52-57.

Ohio University
Athens, Ohio

LES COMPARAISONS DE CAPACITÉ SELON RACE DANS LE RAPPEL IMMÉDIAT DE MATÉRIEL LOGIQUE ET NON-SENS

(Résumé)

On a fait subir les tests de la statue de marbre, du carré de lettres, et les tests auditifs et visuels de la rangée des chiffres à des groupes d'élèves âgés

de douze ans et à des étudiants universitaires, mâles et femelles, représentant les races blanche, japonaise, chinoise, et howaienne. Immédiatement après la présentation on a noté les réponses par écriture.

On donne les moyennes pour les élèves et les étudiants universitaires séparément pour les mâles et les femelles, mais seulement les données pour les groupes âgés de douze ans sont assez adéquates pour justifier plus de procédés statistiques. Les différences entre le rendement de chaque groupe de race-sexe-test et les groupes blancs correspondants sont presque toutes petites et insignifiantes. Comme le montre le Tableau 8, seulement quelque 30 pour cent a plus de deux fois sa σ_{diff} . On trouve aussi que sept des huit différences significantes sont pour les tests de la statue de marbre et du carré de lettres. Dans ces tests il s'agit des facteurs autres que la mémoire lesquels peuvent expliquer les plus grandes différences. Ceci est surtout évident quand on considère les tests où il s'agit plus spécifiquement de mémoire, ceux de la rangée des chiffres, où les différences sont toutes peu élevées et insignifiantes.

Quand on évalue les résultats sur une base de pourcentage et les évaluations de pourcentage pour chaque test et une évaluation composée pour les quatre tests sont comparées selon la race la conclusion citée ci-dessus est plus corroborée.

LOUTTIT

RASSENVERGLEICHUNGEN AN DER FÄHIGKEIT ZUR UNMITTEL- BAREN WIEDERHOLUNG VON LOGISCHEM UND VON SINNLOSEM STOFF

(Referat)

Die Prüfungen genannt "Marmorstatue" (marble statue), "Buchstabe-Quadrat" (letter square), "auditive Zahlspannbreite" (auditory digit span), und "visuelle Zahlspannbreite" (visual memory span) wurden Gruppen bestehend aus zwölfjährigen Mädchen, zwölfjährigen Knaben, und Universitätsstudenten angeboten, in denen "weisse," Japanische, Chinesische, und Hawaiiier Abstammung vertreten waren. Gleich nach der Darbietung wurden die Antworten schriftlich eingetragen.

Die Durchschnittszahlen werden für die Schulkinder und die Studenten und für männliche und weibliche Personen getrennt angegeben. Nur die Daten über die zwölfjährigen Gruppen sind aber für weiteres statistisches Verfahren zureichend. Die Unterschiede in der Arbeit der verschiedenen Rassen-, Geschlechts-, und Testgruppen sind fast durchaus klein und unbedeutend. Wie es die Tabelle 8 zeigt, sind nur etwa 30% von ihnen mehr als zwei mal so gross wie die sigma (σ) ihrer Unterschiede. Es zeigte sich ferner, dass aus den acht bedeutenden Unterschieden sich sieben auf die Marmorstatue Prüfung und die Buchstabe-Rhombusprüfung beziehen. Diese Prüfungen schliessen ausser des Gedächtnisses auch andere Elemente in sich ein welche vielleicht die grösseren Unterschiede erklären. Diese Tatsache wird besonders klar, wenn wir die mehr spezifischen Gedächtnisprüfungen in Betracht ziehen, —namentlich die Prüfungen der Zahlspannbreite, worin alle Unterschiede klein und unbedeutend sind.

Werden die Protocole prozentmässig eingetragen, und werden auf dieser Basis die Zahlen für jede Prüfung und eine Gesamtzahl für die vier Prüfungen mit Bezug auf Rassenunterschiede verglichen, so wird die obenerwähnte Schlussfolgerung bekräftigt.

LOUTTIT

THE FORGETTING OF PLEASANT AND UNPLEASANT EXPERIENCES IN RELATION TO INTELLIGENCE AND ACHIEVEMENT*

From the Psychiatric Clinic and Washington University, St. Louis

H. MELTZER

That differences in forgetting unpleasant impressions are probable manifestations of the influence of moods, sentiments, or complexes—temperamental differences—is suggested by the results of a number of previous investigations.¹ The rôle of such forgetting in the genesis and development of abnormalities or peculiarities of behavior is indicated in many case studies reported in analytic psychology—including the writings of Morton Prince and Janet as well as psychoanalytic literature. Concerning the relationship of human experiencing and forgetting of the unpleasant to other important concepts in psychoanalysis, such as the “pleasure principle” and “repression,” much has been written.² Concerning the significance of such experiencing and forgetting in more normal individuals a number of generalizations are found in psychological literature. These are expressed in terms of correlations with factors which facilitate, inhibit, or are symptomatic of, mental health, educational and vocation success or failure. The nature of these generalizations is indicated in the following brief descriptions.

ARE PESSIMISTIC INDIVIDUALS SUPERIOR IN SCHOLARSHIP BUT INFERIOR IN INTELLIGENCE?

Concerning the 62 beginning students in psychology investigated by Laird in a study called “The Influence of Likes and Dislikes on Memory in Relation to Personality” (4), the author concluded that “those who classed themselves as pessimistic and who were so

*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office June 23, 1930.

¹For results with the use of the association method see Vassar studies of M. F. Washburn and students (11-14). For investigation of personal experiences see Meltzer (8).

²See Morgan (9, pp. 193-217) for an intelligent discussion of this literature, and for a critical review of experimental studies of the Freudian conception of forgetting see Meltzer (7).

classed by their fellow students were invariably above the average, very much so, in scholarship. Their average, however, in intelligence was below the class average." This he interpreted as follows: "Considering their equipment they were getting more out of their work in practically all classes than their more light-hearted fellows. Assuming that life to them was painfully toned the unpleasant was not obliuiscening; or perhaps it was a basal seriousness that encouraged more earnest work before the time comes when the work is done" (5, p. 300). These conclusions Laird considered important enough to include in his book, *The Psychology of Selecting Men* (5). For seeing the need for such facts Laird certainly deserves credit. But one gets the impression in reading his study that this aspect was not carefully investigated. In fact, these rather sweeping conclusions are reported as the result of "an interesting observation made in checking the individual papers." No statistical evidence of any kind is presented. It is, therefore, best considered as a suggestion or tentative hypothesis. The next logical questions to ask, then, are: What light would the results of a quantitative study throw on (a) the constancy of the relationships now described as holding "invariably"? (b) the extent of differences described by the expression, "very much so"?

DO THE EMPTY-HEADED SHOW THE MOST REMINISCENCE?

Ballard's general hypothesis concerning the relationship between reminiscence and intelligence is "that a high degree of reminiscence is indicative of a low grade of intelligence" (2, p. 46), or, as he expressed it elsewhere in the same article, "it is as a rule the empty head that shows the most reminiscence" (2, p. 43). Though his results for normal children do not fully support this hypothesis, the results with a class of 80 Training College students—the group which most nearly corresponds to the one investigated in this study—does confirm it. The materials used by Ballard was ballad poetry. To what extent is his hypothesis applicable for obliuiscence and reminiscence of personal experiences?

MENTAL TEST SCORES AND THE PURSUIT OF PLEASURE

The outstanding feature of individuals who excel in mental tests is their pursuit of pleasure, whereas individuals who excel in school grades possess a greater extent of more desirable qualities. That, in brief, is the conclusion made by three University of Michigan in-

vestigators (1). For institutions who continue to use mental tests as an essential part of requirements for admission they draw a very tragic picture, indeed. Their prediction they express as follows: "Assuming that the experimental results represent an approximately accurate picture of the one who does well in mental tests, we may be very sure that the University would soon become a haven for the slack, the foolish, the playful, the indolent and the slothful. Such a personnel would add very materially to the unpleasant duties of the deans and other officers of administration, would cause an increased percentage of students sent home for various reasons, or an undesirable lowering of the university standards until it would approximate the mental level of the finishing school" (1, p. 277). These are very challenging statements. And, furthermore, are very important to all who are concerned with the problem of selecting or dealing with men. The basis for these predictions is their own rating study of 96 college students and their interpretation of Webb's (15) rating study. To the extent that the ratings are inaccurate or unreliable, the conclusions are, of course, unwarranted. What light will an analysis of the relevant data from the present study—all of which were obtained by more direct and objective measures than ratings—throw on the validity of these very challenging conclusions?

THE PURPOSE OF THE PRESENT STUDY

In short, the purpose of the present study is to investigate the relationships of the remembering and forgetting of pleasant and unpleasant personal experiences to intelligence, as measured by a test score, and to academic grade, as measured by objective tests.

METHOD AND MEASURES

1. *How Facts of Forgetting were Obtained.* On the day following a Christmas vacation 132 individuals—77 men and 55 women—were asked to list and briefly describe all their experiences of the vacation period. They were then instructed to mark *P* before each experience that was pleasant and *U* before every experience which was unpleasant to them. They were then further directed to rate the intensity of each experience.³ These materials were then collected and deposited in the office of the psychology department. Six weeks later, without warning, the foregoing instructions were re-

³The data on the rôle of intensity will be reported separately in a later article.

peated. The experiences listed by each individual at the time of the second recall were then matched with experiences listed at time of first recall. New experiences, i.e., experiences reported on the second recall not given on the first were eliminated from the tabulations on extent of forgetting and treated separately. This procedure yielded for analysis 2231 actual life experiences whose feeling tone was reported by the individuals who did the experiencing, remembering, and forgetting.

2. *Measures and Formulas Used in Working with Facts of Forgetting.* Pleasant, unpleasant, and total experiences reported on the first recall were recorded as P , U , and T , respectively. The symbols used for recording experiences reported on the second recall were P_2 , U_2 , and T_2 . P_n , U_n , and T_n are the symbols used to stand for new experiences; P_f , U_f , and T_f for forgotten experiences.

The forgetting of experiences reported in terms of percentage are self-explanatory.

3. *Measures of Feeling Tone.* To measure the facts of forgetting in relation to the factor of time the following three indices were calculated for each individual:

a. *Pleasant-tone Index.* This measure indicates the direction and extent to which the proportion of pleasant experiences, in percentage, increases or decreases with the passing of a time-interval of six weeks. Mathematically it is simply the difference between the percentage of pleasant experiences reported on the second recall and first recall. The formula thus is: $P_2/T_2 - P_1/T_1$. Plus (+) answers thus indicate a tendency for percentage of pleasant to increase with time; minus (—) answers indicate a tendency to decrease, and zero (0) answers indicate that the percentage of pleasant memories has not changed from first to second recall.

b. *Unpleasant-tone Index.* This measure indicates the direction and extent to which unpleasant feeling tone in terms of percentage of unpleasant experiences changes with increase in time. The formula is: $U_2/T_2 - U_1/T_1$.

c. *P-U Potency Index.* This measure indicates the direction and extent to which, with the passing of time, the tendency to forget the unpleasant is larger or smaller than the tendency to forget the pleasant. In short, it is the difference between the unpleasant-tone index and the pleasant-tone index. Plus (+) differences indicate a greater tendency to forget unpleasant than pleasant experiences;

minus (—) differences indicate a greater tendency to forget the pleasant, and zero (0) differences indicate equal potency.

4. *How Intelligence was Measured.* Intelligence was measured by the Otis Self-Administering Test, Higher Examination, Form A. In calculating the means of superior, mediocre, and inferior intelligence, an Otis score below 47 was considered inferior and a score higher than 55 as superior.

5. *How Achievement was Measured.* Achievement in this study is given a very limited but objectively determined meaning, namely, grade on psychology as determined by six 50-minute objective tests given in well distributed periods of the session. The total scores were transmuted into letter grades used in the institution where this study was made—A, B, C, D, E, and F. For purposes of this study these were regrouped into three divisions as follows: A and B to represent superior achievement; C, mediocre; and D, E, and F, inferior achievements. In the C, or mediocre group, were included all students whose scores were included in $M \pm \frac{1}{2}A.D.$, where M stands for the median and $A.D.$ for the average deviation. Any score below $M - \frac{1}{2}A.D.$ was considered as representing inferior attainment, and any score above $M + \frac{1}{2}A.D.$ as representing superior attainment.

RESULTS

1. *Relation of Experiences Reported on First Recall to Superiority and Inferiority in Intelligence and Academic Grade.* The relationships between the number of pleasant, unpleasant, and total experiences on first recall with intelligence and academic grade will be reported in two ways: (a) correlations calculated by Pearson product-moment formula for factors involved, supplemented by (b) comparable measures of central tendency and variability for individuals classified as inferior, mediocre, and superior in intelligence and grade.

All of the correlations reported in Table 1 are positive. But they are all relatively low. Intelligence test scores correlate highest with pleasant experiences and lowest with unpleasant. Academic grades correlate highest with total experiences reported and lowest with pleasant experiences.

There is very little difference between the mean number of pleasant, unpleasant, or total experiences reported by individuals of superior as compared with either mediocre or inferior intelligence. The out-

TABLE 1
RELATIONSHIPS FOUND ON FIRST RECALL

	<i>r</i> *	Pleasant experiences		<i>r</i>	Unpleasant experiences		<i>r</i>	Total experiences	
		Mean	S.D.		Mean	S.D.		Mean	S.D.
A. Intelligence	.185			.068			.153		
1. Superior		11.7	3.40		7.40	2.00		19.00	4.40
2. Mediocre		11.7	4.80		7.40	3.80		18.60	7.40
3. Inferior		11.7	4.00		7.40	2.80		17.60	5.40
B. Academic grade	.143			.175			.241		
1. Superior		12.33	4.46		7.58	3.22		19.22	6.62
2. Mediocre		10.82	3.74		7.09	2.48		17.36	4.90
3. Inferior		10.07	3.88		6.07	2.42		15.39	5.02

**P.E.*, with grade $\pm .05$ and with intelligence $\pm .08$.

standing difference is the consistently larger variability for mediocre intelligence. The mean number of experiences reported by individuals of superior grade are higher than the number given by mediocre, and that, in turn, is higher than that given by inferior. The variabilities considered, these differences are not as large as they appear to be at first glance.

Do these findings substantiate the conclusions made by the investigators who used rating methods? The results of his own and of Webb's study Adams describes as "pleasingly consistent" (1, p. 264). He further states that "almost without exception they show that the one who excels in school grades possesses desirable qualities to a greater extent than the one who proves facile in mental tests." For purposes of making our comments concrete as well as for comparison with our own findings certain of the results of these investigators will be reported here.

TABLE 2
SELECTED RESULTS FROM RATING STUDIES

Investigator	Trait Rated	Intelligence	Grade
Webb	Mental work on studies	-0.18	0.65
	Mental work on pleasures	0.94	-0.01
	Cheerfulness	0.49	0.24
	Bodily activity for pleasure	0.92	-0.03
Adams	Joyful	-0.08	-0.77
	Memory	0.38	0.77
	Optimism	-0.01	-0.01
	Tendency to worry	-0.12	0.12

Aware of the fallacy of ratings, a committee from the American Council of Education made the following recommendations for their use: "Those who rate individuals should rate only those traits they themselves have had the opportunity to observe and that if they must rate large numbers of students the number of items should not exceed five" (10). In neither of the investigations in question were these recommendations satisfied. Neither were the recent improvements in devices (cf. 6) for rating character applied. In the light of these facts, it seems likely that the subjects rated, not traits, but just so many words—words whose meanings are not uniformly comprehended by psychologists and less likely to have been so comprehended by the subjects. An analysis of the correlations in Table 2 suggests the same conclusion. For example, "optimism" is reported as correlating -0.01 with both intelligence and grade; "joyfulness" $-.08$ with intelligence and $-.77$ with grade; "cheerfulness" 0.49 with intelligence and 0.24 with grade. How many psychologists would agree that the differences between optimism, joyfulness, and cheerfulness are as large as indicated by these correlations?

The results reported in Table 1, on the other hand, are of actual activities recalled by the individuals who did the experiencing. The recall was made, it will be remembered, on the first day after the vacation period in which the activities were experienced. Surely the extent or proportion of pleasant experiences reported on the first recall is a more direct measure of the "pursuit of pleasure" than are the ratings made by friends or acquaintances. At most, the findings suggest some little tendency for more intelligent individuals to be more affectively selective in their experiencing, whereas the more noticeable tendency for individuals who make high grades is just to recall more experiences regardless of the nature of the feeling tone. Even these less sweeping conclusions cannot be made with certainty if it is remembered that individuals who make high academic grades also report substantially more pleasant than unpleasant experiences and that the difference between the correlations of intelligence with P_1 and grade with P_1 is small.

2. *Pleasant and Unpleasant Experiences Forgotten in Relation to Intelligence and Grade.* The extent of the relationships of both intelligence and grade with the measured facts of forgetting for the group treated as a whole is indicated in Table 3.

Comparable measures of central tendency and variability for indi-

TABLE 3
CORRELATIONS WITH MEASURES OF FORGETTING

Percentage of experiences forgotten	Intelligence		Grade	
	<i>r</i>	<i>P.E.</i> _{<i>r</i>}	<i>r</i>	<i>P.E.</i> _{<i>r</i>}
Pleasant	.056	.08	.099	.05
Unpleasant	-.037	.08	-.061	.05
Total	.036	.08	.062	.05

TABLE 4
PERCENTAGE OF EXPERIENCES FORGOTTEN

	<i>N</i>	Pleasant		Unpleasant		Total	
		Mean	<i>S.D.</i>	Mean	<i>S.D.</i>	Mean	<i>S.D.</i>
A. Intelligence	75						
1. Superior	26	48.7	16.08	55.6	21.6	50.8	16.2
2. Mediocre	29	53.3	33.6	65.5	24.0	57.2	15.6
3. Inferior	20	46.7	17.4	61.2	24.0	53.7	15.6
B. Academic grade	130						
1. Superior	45	48.47	17.4	60.2	21.54	51.85	17.05
2. Mediocre	44	45.68	18.0	55.09	23.88	50.8	15.05
3. Inferior	41	43.56	16.8	61.24	27.48	49.7	17.0

viduals classified as inferior, mediocre, or superior in intelligence and grade are reported in Table 4.

In the relationships of the facts of forgetting with intelligence and grade there is again little that suggests any easily distinguishable differentiations. The correlations of unpleasant experiences forgotten are negative with both intelligence and grade—somewhat more negative with grade. All the other correlations are positive, and also low—somewhat higher with grade. The means and *S.D.*'s of experiences forgotten, classified by grade, are more regular and consistent in their direction than are the comparable measures classified in terms of intelligence. Individuals superior in academic grade forgot more pleasant and total experiences than individuals who are mediocre, and these, in turn, forgot more than individuals who made inferior grades. The differences are small. In unpleasant experiences, however, it is the individuals who made inferior grades who forgot most and those who made mediocre grades who forgot least. Considered in terms of intelligence it is the mediocre who did the most forgetting of pleasant and total as well as unpleasant. The superior group forgot more pleasant but less unpleasant and total than the inferior group.

3. *Correlations with Reminiscence.* The relationships found on the first recall and the correlations of the facts of forgetting suggest that the hypothesis of Ballard concerning the relationship of oblivescence and reminiscence with intelligence obtained with the use of ballad poetry as the materials is hardly applicable to personal experiences. At least in part, new experiences—those reported on the second recall though not given on the first—are explained by what Ballard describes as having established with certainty, namely, “no kind of material upon which memory exercises itself is exempt from the operation of reminiscence. Whatever may be remembered may also be forgotten, and what may under certain circumstances be forgotten may, under certain circumstances, be remembered again” (2, p. 28). In this instance, however, almost one-sixth of the subjects did not report any new experiences. And the mean number of new experiences reported is approximately one. The number of new experiences correlates $-.11 \pm .06$ with grade and $.042 \pm .08$ with intelligence. For the recall of personal experiences, then, the results do not substantiate Ballard’s conclusion to the effect that a high percentage of reminiscence is a sign of mental weakness.

4. *Relationships of Feeling Tone Indices.* The amounts or percentages of experiences forgotten represent static measures. They do not indicate the extent and direction to which the proportion of pleasant or unpleasant experiences reported on the first recall increase or decrease with the passing of a time-interval of six weeks. The feeling tone indices previously described are purposed to do just this. How these indices—pleasant tone, unpleasant tone, and *P-U* potency—correlate with intelligence and grade is indicated in Table 5.

The correlations of the feeling tone indices are again all low. The tendency of the percentage of pleasant experiences to increase with the passing of six weeks as measured by the pleasant tone index cor-

TABLE 5
CORRELATIONS WITH FEELING TONE INDICES

Index	Intelligence		Grade	
	<i>r</i>	<i>P.E.</i> _{<i>r</i>}	<i>r</i>	<i>P.E.</i> _{<i>r</i>}
Pleasant tone	-.053	.08	-.14	.05
Unpleasant tone	.005	.08	.13	.05
<i>P-U</i> potency	-.049	.08	-.139	.06

relates negatively with both intelligence and grade—somewhat more negatively with grade. On the other hand, unpleasant tone index correlates positively with both factors—somewhat more so with grade. The *P-U* potency index correlates negatively with both intelligence and grade. To the limited extent that these correlations show a differentiation in the relationship of the forgetting of pleasant and unpleasant experiences with the passing of time, the direction is in accord with Laird's generalization. But the extent of the differences is small and certainly indicate nothing that holds "invariably" and "very much so." As a matter of fact, on *P-U* potency index—which indicates the tendencies in forgetting in a single expression—the difference between the correlations is .09, the $P.E_{diff, r_1-r_2}$ is .094 and the $Diff./P.E_{diff}$ is .95. And this means the chances are 74 in 100 that the true difference is greater than zero (cf. 3, pp. 135, 171). The difference is thus about 25% of what it should be to guarantee reliability.

SUMMARY

Of Procedure

Quantitatively determined facts of remembering and forgetting of pleasant and unpleasant personal experiences were correlated with intelligence and grade in psychology. The findings may be briefly summarized as follows under the captions used in their original presentation.

Of Findings

1. *Relation of Experiences Reported on First Recall to Superiority and Inferiority in Intelligence and Grade.*

a. Intelligence correlates .183 with pleasant experiences reported on first recall, .068 with unpleasant, and .153 with total.

b. Grade correlates .143 with pleasant experiences, .175 with unpleasant, and .241 with total.

c. There is very little difference between the mean number of pleasant, unpleasant, or total experiences reported by individuals of superior as compared with mediocre or inferior intelligence.

d. The mean number of experiences reported by individuals who made superior grades is higher than the mean number reported by individuals who made mediocre grades, and that, in turn, is higher than that reported by individuals who made inferior grades.

The relationships found on the first recall do not substantiate Adams' contention to the effect that the outstanding feature of individuals who excel in mental tests is their pursuit of pleasure whereas individuals who excel in school grades possess a greater extent of desirable qualities. At most, the results suggest some little tendency for more intelligent individuals to be more affectively selective in their experiencing or recall of experiences, whereas the more noticeable tendency for individuals who make high grades is to recall more experiences regardless of the nature of the feeling tone associated with the experiences.

2. *Pleasant and Unpleasant Experiences Forgotten in Relation to Intelligence and Grade.* Correlation of percentage of pleasant, unpleasant, and total experiences forgotten after a six-week interval with intelligence and grade reveal no easily distinguishable differentiations. The correlations are positive with pleasant and total forgotten and negative with unpleasant, but all the correlations are close to zero.

3. *Correlation with Reminiscence.* Ballard's conclusion to the effect that a high percentage of reminiscences is a sign of mental weakness, the basis of which was an investigation wherein ballad poetry was the material used, does not appear to be applicable to personal experiences. Besides the facts previously reported, the number of new experiences—those named on the second recall though not reported on the first—was used as a measure of reminiscence. The number of new experiences was found to correlate $-.11$ with grade and $.042$ with intelligence.

4. *Relationships of Feeling Tone Indices.*

a. The pleasant tone index—a measure indicating the direction and extent to which the percentage of pleasant experiences increases or decreases with the passing of six weeks—correlates $-.053$ with intelligence and $-.14$ with grade.

b. The unpleasant tone index—a measure indicating the direction and extent to which the percentage of unpleasant experiences increases or decreases with the passing of six weeks—correlates $.005$ with intelligence and $.13$ with grade.

c. The *P-U* potency index—a measure which indicates the extent to which the individuals tend to forget the unpleasant more than the pleasant or the pleasant more than the unpleasant—correlates $-.049$ with intelligence and $-.139$ with grade.

To the limited degree that these correlations tend to differentiate

the extent of the oblivescing of the unpleasant of individuals of superior intelligence as contrasted with individuals who made superior grades, the direction indicated is in accord with that described by Laird; namely, the presence of smaller amounts of such oblivescing in individuals who made superior grades. But the extent of the difference, instead of being "very much so" is very small and statistically unreliable ($\text{diff.} = .09$ and $\text{Diff./P.E.}_{\text{diff.}} = .095$).

General Conclusion. The differences found warrant further quantitative study of actual life situations on a larger scale and also suggest the need for the supplementation of such statistical studies by clinical personality studies of at least some of the individuals who did the experiencing, remembering, and forgetting.

REFERENCES

1. ADAMS, H. F., FURNISS, L., & DEBOW, L. A. Personality as revealed by mental test scores and by school grades. *J. Appl. Psychol.*, 1928, **12**, 261-277.
2. BALLARD, P. B. Obliviscence and reminiscence. *Brit. J. Psychol., Monog. Suppl.*, 1913, **1**, No. 2. Pp. vii+82.
3. GARRETT, H. E. Statistics in psychology and education. New York: Longmans, Green, 1926. Pp. 330.
4. LAIRD, D. A. The influence of likes and dislikes on memory in relation to personality. *J. Exper. Psychol.*, 1923, **6**, 294-300.
5. ———. The psychology of selecting men. (2nd ed.) New York: McGraw-Hill, 1927. Pp. 345.
6. MAY, M. A., & HARTSHORNE, H. Recent improvements in devices for rating character. *J. Soc. Psychol.*, 1930, **1**, 66-77.
7. MELTZER, H. The present status of experimental studies of the relationship of feeling to memory. *Psychol. Rev.*, 1930, **37**, 124-139.
8. ———. Sex differences in forgetting pleasant and unpleasant experiences. *J. Abn. & Soc. Psychol.*, 1931, **25**, 450-464.
9. MORGAN, J. J. B. The psychology of abnormal people. New York: Longmans, Green, 1928. Pp. 600.
10. ROBERTSON, D. A. A cooperative experiment in personnel procedure. *School & Soc.*, 1927, **26**, 275-277.
11. WASHBURN, M. F., BAXTER, M. F., & YAMADA, K. Directed recall of pleasant and unpleasant experiences. *Amer. J. Psychol.*, 1917, **28**, 155-157.
12. WASHBURN, M. F., GIANG, F., IVES, M., & POLLACK, M. Memory revivals of emotion as a test of emotional and phlegmatic temperaments. *Amer. J. Psychol.*, 1925, **36**, 456-459.
13. WASHBURN, M. F., HARDING, L., SIMONS, H., & TOMLINSON, D. Further experiments on directed recall as a test of cheerful and depressed temperaments. *Amer. J. Psychol.*, 1925, **36**, 454-456.
14. WASHBURN, M. F., MORGAN, E., & MULL, H. K. An attempt to test moods or temperaments of cheerfulness and depression by directed re-

call of emotionally toned experiences. *Amer. J. Psychol.*, 1919, **30**, 302-304.

15. WEBB, E. Character and intelligence. *Brit. J. Psychol., Monog. Suppl.*, 1915, **1**, No. 3. Pp. ix+99.

*Psychiatric Clinic and Washington University
St. Louis, Missouri*

L'OUBLI DES EXPÉRIENCES AGRÉABLES ET DESAGRÉABLES DANS SON RAPPORT À L'INTELLIGENCE ET À L'ACCOMPLISSEMENT

(Résumé)

Cette étude est un essai de l'application de la méthode quantitative à l'investigation de la signification de l'expérience et de l'oubli des impressions désagréables chez les individus normaux. Le lendemain des vacances et encore six semaines plus tard on a fait à 132 étudiants universitaires, sans avertissement antérieur, noter leurs expériences agréables et désagréables pendant les vacances. Ceci a donné 2231 véritables expériences à analyser. On a déterminé l'intelligence par les résultats d'un test et l'accomplissement par le résultat total de six tests objectifs.

Les relations trouvées dans le premier rappel suggèrent que les individus les plus intelligents tendent quelque peu à choisir plus affectivement dans leurs expériences et leurs souvenirs, tandis que les individus qui ont des résultats élevés montrent une tendance plus marquée à se rappeler plus d'expériences sans égard à la nature du tonus émotif. Toutes les corrélations des expériences oubliées sont presque nulles. Les résultats suggèrent que la conclusion de Ballard à l'égard du rapport étroit de la réminiscence avec la faiblesse mentale n'est pas applicable aux expériences personnelles. Les corrélations des indices du tonus émotif—mesures qui indiquent la direction et l'étendue de l'accroissement ou du décroissement du pourcentage de l'agréable ou du désagréable avec l'écoulement du temps—suggèrent que les individus d'intelligence supérieure oublient plus le désagréable que les individus d'accomplissement supérieur. L'étendue de la différence, au lieu d'être selon Laird "beaucoup," est très petite et ne possède pas une valeur statistique.

MELTZER

DAS VERGESSEN VON ANGENEHMEN UND ANGENEHMEN ERFAHRUNGEN IN BEZUG AUF INTELLIGENZ UND ERFOLG

(Referat)

In dieser Untersuchung versuchte man, bei der Erforschung der Bedeutung des Erlebens und des Vergessens von unangenehmen Eindrücken bei normalen Individuen-quantitative Methoden anzuwenden. Den nächsten Tag nach Ferien, und wieder sechs Wochen später, wurden 132 Studenten ohne Kündigung ersucht, ihre angenehmen und unangenehmen Erfahrungen aus den Ferien zu notieren. Dieses Verfahren bot 2231 echte Erlebnisse zur Analyse dar. Es wurde die Intelligenz durch eine Testabschätzung und Erfolg durch die Gesamtaberschätzung an sechs objektiven Tests ermittelt.

Die Verhältnisse, welche sich in Bezug auf die erste Erinnerung erwiesen, scheinen anzudeuten, dass die intelligenteren Individuen dahin neigten, affectiv in ihrem Erleben oder Erinnern mehr wählerisch zu sein, während bei den Individuen die in ihren Kursen guten Erfolg erzielten (make high grades) sich eher die Neigung, sich an eine grössere Anzahl von Erlebnissen, ohne Bezug auf deren Gefühlston, zu erinnern, merken lies. Alle Korrelationen über die vergessenen Erlebnisse gruppieren sich um zero herum. Die Ergebnisse deuten an, dass Ballard's Schluss über das nahe Verhältnis zwischen Reminiszenz und geistige Schwäche sich nicht auf persönliche Erfahrungen anwenden lässt. Die Korrelationen von Anzeigern des Gefühlstons (feeling-tone indices) —d.h., von Messungen, welche die Richtung und den Grad des prozentualen Abnehmens und Zunehmens des Angenehmen und Unangenehmen mit Lauf der Zeit angeben, —scheinen anzudeuten, dass sich bei Personen von überlegener Intelligenz in Vergleich mit Personen von überlegenem Erfolg mehr Obliviszenz unangenehmer Erfahrungen zeigt. Der Umfang des Unterschieds, anstatt, wie nach Laird's Beschreibung, "sehr gross" zu sein, ist sehr klein, und statistisch unzuverlässig.

MELTZER

SHORT ARTICLES AND NOTES

THE MEASUREMENT OF CHANGE IN SOCIAL ATTITUDE*

L. L. THURSTONE

The experiment to be described in this paper was set up in order to ascertain whether the effect of a single motion picture on the social attitudes of school children could be measured by an attitude scale in the statement form. The plan was, in brief, to let the school children of one town see a film favorable to Chinese culture and in a second town the school children were shown a film that has been criticized as unfriendly to the Chinese. Before and after seeing the film in their respective towns, the children filled in a statement scale about the Chinese. The results show that the attitudes of the children were changed in opposite directions in the two towns, thus demonstrating the effect of the films as well as verifying the methods used.

The two towns selected for the experiments were Geneva, Illinois, and West Chicago, Illinois. The relevant facts about these two towns are given in Table 1. The film "Son of the Gods" has been considered rather friendly in its interpretation of Chinese culture. It was shown in Geneva, Illinois. One week before and the day after seeing this picture, the children filled in a statement scale about the Chinese. This attitude scale was constructed by the method of equal appearing intervals. The scale value of each statement is recorded for each opinion in the list. It was constructed by the pooled judgments of 30 subjects. Since the detailed methods of constructing an attitude scale have been described elsewhere (1), they will not be repeated here.

In Figure 1 we have the frequency distribution of attitudes of these children to the Chinese as determined on May 19, 1930, before they saw the picture. The picture was shown in the local theater by special arrangement on May 26, 1930. On the morning after seeing the picture, the children again filled in the statement scale about the Chinese. They had been told

*This is one of a series of experimental studies on the effect of motion pictures on the social attitudes of children. These studies were made possible by a grant from the Payne Fund. The author wishes to acknowledge the assistance of Professor W. W. Charters of Ohio State University, Mr. C. C. Byerly, School Superintendent at West Chicago, and Mr. H. M. Coultrap, School Superintendent at Geneva, Illinois. Miss Ruth C. Peterson conducted the experiments. We wish to acknowledge the assistance of the First National Picture Distributing Corporation, Mr. Paul Polka of Maywood, and Mr. Rubens of Balaban and Katz in making special arrangements about the film, "Son of the Gods;" and of Paramount Famous Lasky Corporation in the arrangements for the film "Welcome Danger."

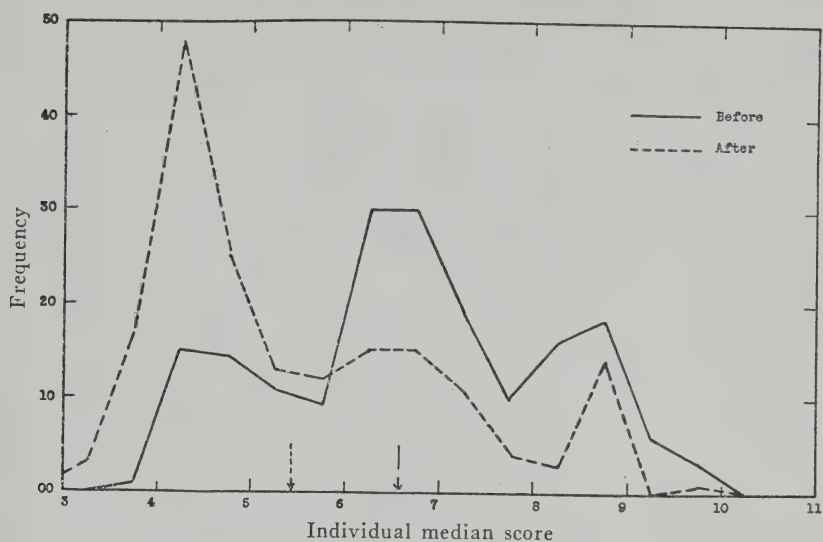


FIGURE 1

FREQUENCY DISTRIBUTION OF SCORES OF 182 HIGH-SCHOOL CHILDREN IN GENEVA, ILLINOIS, ON A SCALE OF ATTITUDE TOWARD THE CHINESE BEFORE AND AFTER SEEING THE FILM "SON OF THE GODS"

previously that the scale was to be filled in twice several days apart. In Figure 1 we have shown also the frequency distribution of their attitudes toward the Chinese after seeing the film. The shift in attitude as a result of the picture is striking. The picture evidently made the children more friendly toward the Chinese. The statistical facts about the two frequency distributions are as follows:

Number of children	182
Mean attitude score (before)	6.63
Mean attitude score (after)	5.45
Standard deviation of scores (before)	1.46
Standard deviation of scores (after)	1.54
Correlation between the two sets of scores	+ .57
Ratio of the difference to the probable error of the difference	16.98

We are undoubtedly justified in concluding that the film "Son of the Gods" has the effect of making the children more friendly toward the Chinese. It is of some interest to establish that the effect of a single film on the social attitudes of children can be measured by a statement scale.

TABLE 1

Town in Illinois	Motion picture given	Theater	Size of town (1920 census)	Size of school	No. in exp. group	Grades	Date of		Date of	
							first scale	motion picture	second scale	
Geneva West Chicago	"Son of the Gods"	Fargo	2803	230	182	9-12	5/19	5/26	5/27	
	"Welcome Danger"	School	2594	225	172	9-12	5/19	5/26	5/27	

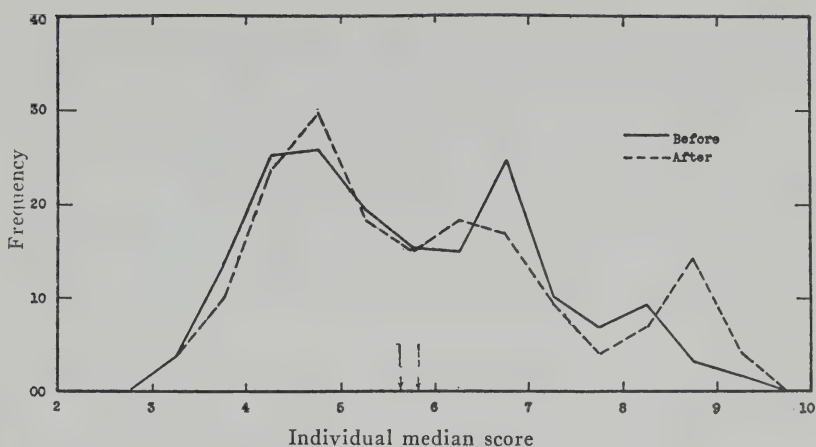


FIGURE 2

FREQUENCY DISTRIBUTION OF SCORES OF 172 HIGH-SCHOOL CHILDREN IN WEST CHICAGO, ILLINOIS, ON A SCALE OF ATTITUDE TOWARD THE CHINESE BEFORE AND AFTER SEEING THE FILM "WELCOME DANGER"

It is more difficult to find films that are antagonistic to foreign nationalities than to find films that are, on the whole, friendly toward foreign cultures. The film "Welcome Danger" has been criticized by the Chinese for its unfriendly manner of dealing with the Chinese and it was therefore selected for this experiment. The film was shown in West Chicago. One week before and the day after seeing the picture, the children filled in the same scale about the Chinese that was used in Geneva. The two frequency distributions, before and after, are shown in Figure 2. Here there is evident a slight shift in the opposite direction. The statistical facts about the two frequency distributions are as follows:

Number of children	172
Mean attitude (before)	5.66
Mean attitude (after)	5.81
Standard deviation (before)	1.42
Standard deviation (after)	1.56
Correlation between two sets of attitude scores	+ .58
Ratio of the difference to the probable error of the difference	2.22

It is evident that the film "Welcome Danger" made the children somewhat more unfriendly toward the Chinese, although the effect of this film against the Chinese was by no means so marked as the effect of the previous film favorable to the Chinese. It would be interesting to compare the

relative ease of moving an audience in the favorable and in the unfavorable directions about a foreign country or race. Unfortunately, that cannot be done experimentally unless the appeal of the two stimuli can be equated by criteria other than their measured effect on the audience.

The present experiments show experimentally that a single film has a measurable effect on the international attitudes of school children and that these effects can be measured by a statement scale.

The statement scale about the Chinese contained the following instructions and opinions. The scale value is recorded for each opinion in the following list. The attitude score is the median scale value of all the opinions endorsed by the subject. The scale values are not printed on the blanks given to the subjects.

This is a study of attitudes toward the Chinese. On the other side of this page you will find a number of statements expressing different attitudes toward the Chinese.

✓ Put a check mark if you agree with the statement.

✗ Put a cross if you disagree with the statement.

Try to indicate either agreement or disagreement for each statement. If you simply cannot decide about a statement you may mark it with a question mark.

This is not an examination. There are no right or wrong answers to these statements. This is simply a study of people's attitudes toward the Chinese. Please indicate your own convictions by a check mark when you agree and by a cross when you disagree.

✓ Put a check mark if you agree with the statement.

✗ Put a cross if you disagree with the statement.

Scale
value

- | | |
|------|--|
| 6.5 | 1. I have no particular love nor hate for the Chinese. |
| 10.1 | 2. I dislike the Chinese more every time I see one. |
| 4.7 | 3. The Chinese are pretty decent. |
| 7.2 | 4. Some Chinese traits are admirable but on the whole I don't like them. |
| .5 | 5. The Chinese are superior to all other races. |
| 8.7 | 6. The Chinese as part of the yellow race are inferior to the white race. |
| 3.5 | 7. I like the Chinese. |
| 2.8 | 8. The more I know about the Chinese the better I like them. |
| 11.0 | 9. The Chinese are aptly described by the term "yellow devils." |
| 1.8 | 10. The high class Chinese are superior to us. |
| 5.2 | 11. The Chinese are different but not inferior. |
| 11.5 | 12. I hate the Chinese. |
| 4.1 | 13. Chinese parents are unusually devoted to their children. |
| 7.7 | 14. Although I respect some of their qualities, I could never consider a Chinese as my friend. |
| 1.2 | 15. I would rather live in China than any other place in the world. |

- 9.7 16. There are no refined nor cultured Chinese.
- 6.0 17. The Chinese are no better and no worse than any other people.
- 8.4 18. I think Chinese should be kept out of the United States.
- 2.2 19. I consider it a privilege to associate with Chinese people.
- 10.6 20. The Chinese are inferior in every way.
- 9.4 21. I don't see how anyone could ever like the Chinese.
- 3.0 22. Chinese have a very high sense of honor.
- 8.6 23. I have no desire to know any Chinese.
- 1.4 24. Chinese people have a refinement and depth of feeling that you don't find anywhere else.
- 9.8 25. There is nothing about the Chinese that I like or admire.
- 3.9 26. I'd like to know more Chinese people.

REFERENCE

1. THURSTONE, L. L., & CHAVE, E. J. The measurement of attitude. Chicago: Univ. Chicago Press, 1929. Pp. 96.

University of Chicago
Chicago, Illinois

EMPATHY AND THE ACTOR'S EMOTION¹

JOHN T. METCALF

There is perhaps no field of art in which empathy plays a more prominent rôle than in the art of the theatre. When conditions are favorable the response of the audience is empathic to a degree that seems impossible elsewhere. It is the writer's belief that empathy also plays an important part in the consciousness of the actor, at least in the early stages of the study of a rôle.

The question as to what the actor really does in feeling himself into a rôle, especially in its emotional aspects, has been variously answered. In general, two sharply opposed views have been held on this subject. One is that the actor is an individual of such facile emotion, so unstable and so lacking in individuality, that he can assume readily the characteristics of another person. Thus, without any real personality of his own, he becomes successively different persons, much as the psychoneurotic vacillates between his alternating personalities. His emotion in playing a part is, according to this view, as real as any emotion he ever experiences.

The other extreme view found its most consistent expression in the writings of Diderot (3), who held that the actor, if he is to do his work effectively, must remain perfectly cold even while playing the most impass-

¹Read in the Symposium on Esthetics at the Ninth International Congress of Psychology, New Haven, Conn., September, 1929.

sioned scenes. Diderot even went so far as to maintain that sensibility, far from being an advantage to the actor, was actually a serious handicap.

It seems to the writer that both of these views are mistaken ones, and that each interprets the situation too simply. To say that the actor merges his own personality fully into that of the fictitious character he is representing is absurd. If this were so, then two actors of equal ability would, in playing the same rôle, become exactly alike, and we know that this is not the case. The writer has seen six different actors, all great, play the rôle of Hamlet; and all six of the Hamlets were convincing as Hamlet, but they were all different. Moreover, a fine actor is nearly always recognizable as himself whatever the rôle in which he appears, just as the style of a writer may be readily identified in literary productions dealing with the most varied subjects.

As for the other view, that acting is entirely a matter of cold, detached, intellectual calculation, it seems to be almost equally far from the truth. The great French actor, Constant Coquelin (2), agreed in the main with Diderot's arraignment of the opposing view, but maintained, nevertheless, that sensibility in the actor was a valuable, indeed an indispensable attribute. He held that in acting the mind of the actor was characterized by a kind of double consciousness, one part of it being devoted to the character portrayed, the other part maintaining a watchful and critical attitude on the part of the actor's own real self. This double consciousness has been attested by so many prominent actors and actresses (1) that it seems to the writer to contain the essence of the true explanation.

If acting were solely a matter of emotion or of intellectual calculation this double consciousness could not appear. What does make it possible is the play of imagination, a quality which is probably more important for acting than either of the others. If the actor cannot vividly represent to himself in imagination the mental attitude of the character he is to portray, it is hopeless for him to try to represent it to other people. Imagining a given mental state tends to stimulate the motor responses appropriate to that state, and these, once produced, are controlled, modified, selected, and developed through rehearsal in the interest of the art of the theatre.

On the stage it is, of course, the overt responses that count, for they are the only ones the audience sees or hears. Internal responses are of importance only insofar as they aid in producing the overt ones more effectively. Now if it were a case of displaying a real emotion, it would, of course, be necessary to arouse all of the motor responses, both external and internal. The reaction of the actor in playing an emotional scene is not a real emotion; it is rather a complex of responses that *represent* a real emotion. The writer suggests that psychologically this complex differs from a real emotion in that some of its elements—chiefly the organic processes—are imaginal instead of sensational. In distinguishing the kinaesthetic image from the kinaesthetic sensation, Titchener (4) writes:

" actual movement always brings into play more muscles than are necessary, while ideal movement is confined to the precise group of muscles concerned. You will notice the difference at once—provided that you have kinaesthetic images—if you compare an actual nod of the head with the mental nod that signifies assent to an argument, or the actual frown and wrinkling of the forehead with the mental frown that signifies perplexity. The sensed nod and frown are coarse and rough in outline; the imagined nod and frown are cleanly and delicately traced." Now it seems to the writer that a similar difference holds as between organic image and organic sensation. The imagined emotion, in its organic constituents, differs from the real one in that it does not call into play all the internal reactions that are involved in the latter. If for any reason they are all called into play, the actor's emotion becomes a real emotion.

The actor's emotion must not cross the tenuous dividing line between the imagined and the real, for that would be fatal to the empathic response of the audience. It would make all the difference between *playing* a scene, which is to be desired, and *making* a scene, which is to be avoided. Real emotions are out of place on the stage. If, for instance, an actor who is representing fear suddenly becomes afflicted with stage-fright, the real fear, far from improving the performance, actually spoils it. The audience wants to have its imagination stimulated, and welcomes the actor whose playing has this effect. The mutual attitude of actor and audience is, or should be, similar to that of children when they say, "Let's pretend." If the actor gives way to real emotion, the audience, which is very quick to sense the fact, feels that the agreement has been broken, and instead of the delightful play of imagination they experience constraint, embarrassment, distress, or possibly malicious amusement.

The actor's consciousness, then, is a dual one, a thing that imagination makes possible. The actor is being himself with reference to a real situation and is at the same time acting as a fictitious character with reference to an imaginary situation. These two situations, the real and the imaginary, may be, and of course normally are, two aspects of the same total situation. We might designate the actor's response to the first as the real reaction and his response to the second as the play reaction. There is a close relationship and a constant interplay between the two, and just here we find the basis of the strongly affective quality of the actor's own consciousness. Pleasantness results when there is harmony between the two responses, that is, when the play response is effective and accomplishes what the real reaction tendency is also set to accomplish so that the two work together. When, however, the play response is ineffectual and comes into conflict with the reaction tendencies of the real response, the result is unpleasantness. This is simply to say that when the actor is satisfied with his playing the experience is pleasant, when he feels that he is bungling, the effect is unpleasant. The affective quality of the actor's consciousness,

then, depends not as much upon the emotions he portrays as upon the degree of effectiveness with which he feels he portrays them.

REFERENCES

1. ARCHER, W. *Masks or faces?* London: Longmans, Green, 1888.
2. COQUELIN, C. *Art and the actor.* (Trans. by A. L. Alger.) New York: Pub. Dramat. Mus., Columbia Univ., 1915.
3. DIDEROT, D. *Le paradoxe sur le comedien.* (Trans. by W. H. Pollock.) London: Chatto & Windus, 1883.
4. TITCHENER, E. B. *Experimental psychology of the thought processes.* New York: Macmillan, 1909.

University of Vermont
Burlington, Vermont

CZECHOSLOVAK PSYCHOLOGY OF TESTING

D. D. DROBA

The purpose of this article is to inform American professional psychologists about some developments in Czechoslovak psychology. Very few know what has been accomplished in this field in the new republic of Czechoslovakia. This condition is largely due to the fact that no article or chapter on the above topic, so far as I am aware, has as yet been published in the American psychological journals or books. The only informational review of Czechoslovak psychology ever published in the English language appeared in the *Scandinavian Scientific Review*, 1924, 3, Nos. 3-4, Oslo, Norway. The review was written by F. Šeracký and J. Kozák, Professors of Psychology and Philosophy, respectively, at the Charles University of Praha, and is a very brief and general one covering only three printed pages.

In the past decade Czechoslovak psychology has undergone considerable development. Several courses in psychology are now given at each of the following universities: Charles University at Praha, Masaryk University at Brno, and Komenský University at Bratislava. Books and articles on psychology have been written and published in Czech language. Psychology is now applied to vocational guidance and selection, to the needs of developing the army, to improving efficiency of work in factories, to education, and so on. Institutes have been founded either entirely or partly for promoting psychological science. To cover all this material is beyond the scope of this article. Only one phase of Czechoslovak psychology will be selected and described on the following pages: the Czechoslovak psychology of testing.¹

¹I am indebted to Dr. F. Šeracký, Professor of Experimental Psychology at the Charles University of Praha, for the information included in this review.

1. INSTITUTE FOR APPLIED PSYCHOLOGY

The most important single factor in promoting psychological testing in Czechoslovakia is the Institute for Applied Psychology at Praha. It was organized in March of 1921 under the direction of Dr. Ružka as a department of the Masaryk Academy of Work. Members of the Institute soon went abroad to learn about the methods used elsewhere. Study tours were made to Dresden, Berlin, Essen, Hamburg, Paris, and London. Upon their return, foreign testing instruments were adapted to home conditions and used in a testing program. Two hundred and forty-five boys and 39 girls were examined in the interest of vocational guidance, and 67 mechanical and electrotechnical employees to see if test results harmonize with success in work.

In the two years that followed an increasing number of persons were given psychological tests. In 1922 a total of 1303 persons were examined, including 750 boys and 140 girls for vocational guidance purposes, and 513 pupils of the State Technical School at Praha. The next year 1038 boys and 325 girls went through the vocational guidance routine. The Department of National Defense became interested in the work of the Institute and requested the testing of 1200 soldiers and 68 pupils of a supply school. The Department of Railways requested that 400 applicants should be given aptitude tests.

In the year of 1924 the work of the Institute somewhat changed. The vocational guidance section became separate and the Institute concentrated on vocational selection. Dr. Forster, the new head, outlined the purposes of the Institute as follows: (*a*) the examination of physiological functions, (*b*) the examination of mental abilities, (*c*) the study of physical and mental efficiency, and (*d*) the study of correlations between mental and physiological abilities. Two thousand nine hundred and fifty-three persons were examined from 1924 up to the end of 1926, including university students, employees, musical college students, soldiers, railroad applicants, drivers, and aviators.

In the period between the years of 1927 and 1929 inclusive, under the direction of Dr. Šeracký, a larger number of middle school and university students were given tests. At this time the interest of the textile industries was aroused and over 4000 employees were examined. The total number of persons examined in these three years was 9438.

The methods used at the Institute may be divided into four types. The first type used is an apparatus constructed by Dr. Forster for the purpose of measuring the distribution of attention. The apparatus consists of a glass board divided into three fields. In each of the fields are found asymmetrically arranged small circles that can be lighted at will by lights of various colors. The subject has in front of him five keys of which three center ones correspond to the three fields on the board. If a white light appears in one of the fields the subject is instructed to press the key on the left end

and at the same time to press the key corresponding to the field in which the light appeared. If two green or red lights appear in two fields, the subject is asked to press the corresponding keys. If no lights appear the key on the right end is to be pressed. There are 13 such situations. A kymograph registers the accuracy as well as the speed of the reactions. Scoring is based on a scale of five grades according to the speed of reaction. Reaction-time below 2.6 sec. is given the first grade, 2.61 to 2.96 the second grade, 2.97 to 3.45 the third grade, 3.46-4.14 the fourth grade, and above 4.15 sec. the fifth grade.

The second type of method is Henry's dynamograph. Instructions are to squeeze a bulb filled with mercury with the right hand as hard as possible and hold it as long as possible. The mercury expelled from the bulb will rise in the glass tube and will fall proportionately to the release of the bulb. The bulb is squeezed until the mercury falls below half of the maximum height attained. An attached kymograph records the results.

The third type of testing is the analysis of spatial relations. Geometric figures are presented to the subject on a cross-section paper by a tachistoscope. The task is to reproduce the figures on a similar cross-section paper. The score is the number of errors made.

The fourth type of testing is the use of intelligence tests. Army Alpha and Beta tests were adapted for use in motormen examinations. A mechanical calculation test was substituted for the second test of the Army Alpha series. The second test of Army Beta was used in place of the fourth test of the Army Alpha. In the eighth test of the Army Alpha questions about machines were inserted. The total number of scores is 188 instead of 212. From the Army Beta especially Tests 2, 5, and 7 are used. Terman's revision of Binet-Simon test is the third intelligence test used by the Institute.

The average intelligence of 667 motormen applicants in 1926 was found to be 61.97 points. In 1927, 461 motormen candidates obtained an average of 69.05 points, which is somewhat higher than in the previous year. Candidates with higher education, 153 in number, attained an average of 79.66 points, while 306 candidates with lower education scored only an average of 66.29 points.

Musical ability was tested by the Seashore tests. Administration was directed chiefly by Dr. Stavel. The pupils of the State Musical College at Praha were used for subjects. It was found that a 75-per-cent agreement exists between the school grades and the scores on the test. This would indicate that in 75% of the cases it is possible to predict the pupil's success in college by his score on the music test.

Intelligence tests were also used among middle school and university students for vocational guidance purposes. Middle school graduates were tested in order to see if they should be advised to enter the university or not. German as well as Czech students were examined and compared. Also differences between Gymnasium graduates and "Real" school graduates

were ascertained. Sex differences were found, correlations were calculated between the intelligence scores and school achievement.

2. ARMY TESTING

Impetus to army testing came from General Dr. L. Fischer, Director of the Army Health Service. During a visit in U. S. A. he realized some of the advantages of army testing and upon his return requested the services of the Institute for Applied Psychology. Army pilots were among the first to be tested. Later the Department of National Defense established a special Psychological Service Station at the Army Institute for Aeronautics under the direction of Dr. V. Forster, Director of the Institute for Applied Psychology.

Dr. Forster has introduced into the Army Institute methods used at the Institute for Applied Psychology described above. Tests to be used in the army were, however, separately standardized to suit the needs of the soldiers. The Army Alpha and Beta were used, also some of the Thorndike tests. For measuring higher intelligence Forster has constructed a test of judging ability.

The minimum requirement for airmen was set at 92 points for the rank-and-file men, and 105 points for officers. Applicants were also tested for distribution of attention by the apparatus already described. The best score obtained was 1.19 seconds, the lowest score was more than 8 seconds, and the average score was 3.4 seconds.

The period of 1927-28 marks the beginning of testing army drivers. It was discovered that people with a driver's certificate from civil life are in intellectual ability and aptitude far below the requirements set for such positions. This circumstance became an impetus for an extension of the use of psychological methods in determining the mental ability of applicants.

Intelligence tests were used for officers as a part of entrance examinations into advanced study courses. According to the results obtained, only 12% of the officers do not satisfy the intellectual requirements. However, testing officers is not very urgent because the army authorities admit to officers' academies only such middle school graduates who have proved to be good students. Testing under-officer candidates is more urgent because selection is based largely on army examinations, and this is to be done within one month after the recruit is enrolled. Correlations were calculated between the psychological scores and success in schools for under-officers. The following correlations were obtained in the different departments: .77, .70, .62, .57, .55. It is, of course, understood that psychological tests measure only one or two types of abilities. There are other abilities prescribed by the army authorities. The findings of psychologists may sometimes conflict with the findings of army men. So it was found that about 20% of the intellectually able men were not acceptable to the army authorities. However, this did not affect the value of psychological testing methods.

3. TESTING AT THE ELECTRIC WORKS OF PRAHA

In 1926 there were 4187 street-car employees in Praha among whom were 2423 conductors. Every day 674 cars pass on 20 tracks 58 miles long. About 516,000 persons were carried daily on the cars. These figures are exclusive of the thousands of people transported by the bus lines.

Testing of the employees of the electric companies was started at the Institute for Applied Psychology. Later examinations were carried on independently of the Institute under the direction of Engineer M. V. Zavorka, and under the supervision of the city of Praha. Since the end of the year of 1926 Dr. Váňa has also worked there as a psychologist. From the beginning up to the end of June of 1927 as many as 2559 persons were examined, the majority of whom were applicants for positions as street-car conductors, though more than 300 were applicants for positions as autobus conductors.

The applicants are tested for acuity of vision and hearing, stereoscopic vision, ability to perceive colors, and scotopic vision. Simple reaction-time tests were also given. The average time for visual stimuli was 19 hundredths of a second, and for auditory stimuli 16 hundredths of a second. Distribution of attention was tested by the Forster apparatus. Henry's dynamometer was also used and the emotions were measured by a galvanometer. Intelligence was measured by the Army Alpha test.

It was found that optimal intelligence quotient for conductors falls in the range from 86 to 90 points. Results showed also considerable agreement between intelligence test scores and scores on the distribution of attention. The correlation coefficient between the reaction-time and Alpha scores was found to be .41.

In order to show the significance of psychological tests, certain comparisons were made. Classifications of applicants based on psychological examinations were compared with classifications based on success in a four-weeks' training course. Psychological classification, whether according to intelligence quotient or based upon all the tests, consists of five grades, designated by numbers from 1 to 5, 1 being the best grade. If we take only those who received grades 1 to 3, based merely on IQ, of a group of 351 applicants tested in the first half of 1926, we find 77.1% agreement between the two classifications. In another group of 390 applicants examined between November of 1926 and February of 1927 the agreement between the two classifications is 93%. If results of all the psychological tests used are the basis of classification then the classification agrees with the training course findings in the first group in 91% of the cases, and in the second group in 95.2% of the cases.

4. TESTING AT THE BATA SHOE FACTORY

The purpose of psychological testing at the Baťa shoe factory is the scientific selection and assignment of applicants to the work for which they

are best fitted mentally. This plan is not merely in the interest of the company but also for the benefit of the employees who are given an opportunity from the very first day to develop their abilities in a job well-adapted to their mental equipment.

Baťa's psychological laboratory has still another purpose. Working in closest cooperation with the Department of Research, psychological research was started on the various abilities which make for efficiency of work in the factory. The first study was done in the sawing factory. Attention was measured by Bourdon's examination, speed and accuracy of movements by Ruffer's apparatus, agility of fingers by a simple device, intelligence by two tests taken from the American Army examinations. In all examinations both time and quality scores were used. Tests were evaluated by comparing results with actual efficiency in work with employees employed for a longer period of time. Workers having the more complicated and responsible positions scored much higher on the psychological tests than workers employed in easier and subordinate positions. For this reason the psychological examinations could be used with a good measure of confidence for selection of new employees.

Selection of boys, ages 14 to 16, for the company's shoe academy was made by the psychological laboratory. Within a short period of time about 400 boys were examined with several psychological tests, mostly to measure intelligence. Out of the 400 boys, of whom some were outsiders as well as others already employed in the factory, 80 were selected for the academy.

Psychological tests were given also to office workers employed in the factory. Altogether, six tests were used to measure such functions as ability to concentrate, ability to calculate, and ability to write business letters. The composite scores of all the six tests were found to agree fairly closely with the actual efficiency in work. This circumstance made it possible to use the tests as part of the admission examination for new applicants.

5. VOCATIONAL GUIDANCE TESTING

The beginning of vocational guidance work in Czechoslovakia dates back to the year of 1919. The Adolescent Club established a department for building up vocational guidance. Mr. Weigner, director of the Technological Museum, Dr. Kučera, and Mr. Dolenský have worked there. When the Adolescent Club dissolved in 1921 the vocational guidance section joined the Masaryk Academy of Work and in May of that year the first Vocational Guidance Bureau was opened at Praha. The Bureau was headed by Mrs. Marie Zandtová and the examinations were carried out by the Institute for Applied Psychology. The first set of tests was prepared by Professor Lörsch in accordance with the advice obtained from Professor Blumenfeld of Dresden.

In 1923 the Bureau became quite independent from the Institute and was

established as the Central Vocational Guidance Bureau for the whole republic. A Board of Trustees, headed by Dr. Gustav Winkler, was formed, consisting of the representatives of the Department of Education, Department of Health and Physical Education, Department of Social Welfare, Chamber of Commerce, and other public institutions. Dr. F. Šeracký was made the supervising psychologist of the Psychological Department of the Bureau. Testing was carried out by Novák.

The test material was first evaluated by the use of a large number of subjects. In four years' time over 200 tests and examinations were used. Many of them were discarded as unreliable. Only satisfactory and reliable tests were recommended to other vocational guidance bureaus which were soon opened in various sections of the country. Tests were delivered to these sectional bureaus all ready for use. Apparatus was expensive and, therefore, was used but little.

Test results were always compared with school judgments. Teachers' estimates were found to be most useful in the case of general intelligence. The coefficient of correlation between the intelligence test scores and teachers' judgments reached as high as .81. Correlations between teachers' estimates and aptitude test results are lower.

The Bureau was using a questionnaire to be filled by the employer after the employee has worked with him at least one year for the purpose of checking up on the results. However, all cases could not be followed up. By the end of 1924, 416 out of 1364 cases tested at the Bureau in 1923 were followed up. Out of this number there were only 58 cases (14%) in which the advice of the Bureau was not followed. Those who did not follow the guidance stated that they could not find suitable positions. In some cases failure was probably due to the attitude of the parents, but many more did not succeed in work because of moral laxity. Out of 358 who followed the guidance, 27, that is 7.54% did not succeed because of moral reasons. Out of 58 who did not follow the advice, 8, that is 13.7%, have morally failed. Fifty per cent of those who have refused to follow the Bureau's advice have failed, while 39.88% of those who accepted the advice have succeeded with excellent results.

Another check upon the procedure used is re-examination after the lapse of one year. Some of the tests do not show any change in scores, others show some change, usually for the better especially if the subject has attended some school during the intervening year. This is true for tests of intelligence, tests of technical abilities, skill, and other abilities. However, tests of attention and tests in which attention forms an important part do not show consistent results.

Altogether there are 21 vocational guidance bureaus in Czechoslovakia under the direction of the Central Bureau at Praha. Bureaus at Pardubice, under the direction of Professor Mladek, and at Brno, under the direction of Professor Chlup, are using locally developed methods. Other bureaus

are dependent in this respect upon the central bureau. Some of the more significant bureaus besides the above two are at Plzeň, Moravská Ostrava, and Valašské Meziříčí. In the Central Bureau at Praha between the years of 1924 and 1927 a total of 7149 persons were examined.

According to this historical sketch the major factor in the Czechoslovak psychology of testing was the Institute for Applied Psychology at Praha, which has developed rapidly during the eight years of its existence. The American army tests in an adapted form were used in the army. An extensive testing program was installed at the electric works of Praha. Baťa has introduced psychological methods to discover abilities which make for efficiency of work in his factory. Finally, a chain of vocational guidance bureaus, using psychological tests, was established throughout Czechoslovakia.

Ohio State University
Columbus, Ohio

A COMPARISON OF LINGUISTIC AND NON-LINGUISTIC ABILITY

HERBERT MOORE

It is usually taken for granted that linguistic ability is that whose strength can be measured by tests involving the language factor, and non-linguistic ability that whose strength can be measured by tests concerned with manipulation. The contrast is often referred to as ability to deal with ideas compared with ability to deal with things. The discrepancy in the results from tests of the two abilities lends support to the thesis that they are two separate functions, with no necessary correlation between the two. This assumption is based on the everyday observation of those who are strong in the one and weak in the other, and very rarely efficient in both, and on the constant evidences of contrast between classes: the mechanic who can do everything with a set of tools and very little with his mother tongue is contrasted with the professor who can find the appropriate word for any occasion and yet can scarcely drive a nail straight; the boy prodigy who lingers over books beyond his age and can scarcely be induced to *do* anything is contrasted with the promising inventor who is constantly showing some new device of his creation; the student in Arts is contrasted with the student in Technology; and in each case the implication is that the one class has much superior strength in one ability than the other. Thorndike voices the same note in his observation that, "We know that, taking people as we find them, the ability measured by verbal tests is not the same as the ability measured by non-verbal tests" (9, p. 126).

Ever since the birth of performance tests, despite the fact that they were introduced as a substitute for linguistic tests in the interests of those with a

language handicap, the contrast has been drawn between the results they produce and the results coming from linguistic tests. The consequence of repeated manifestations of that discrepancy is the inference on the part of many that they are measures of two different functions, between which there is no necessary correlation. Miss Mateer finds this discrepancy so great that she doubts their results comparable (5, p. 186). F. L. Wells thinks it a mistake to accept the results from performance tests as of equal value with the results from the Stanford-Binet. "The material is not the same, nor are the same traits being measured." At their best they are "traits of comparable importance in life adjustment" (10, pp. 120-121).

On the other hand, there are some who claim that linguistic and non-linguistic efficiency are both indicators of the same common function, and both different aspects of the force responsible for adequate response. Spearman accepts a common factor responsible for both except in rare cases, and used the tetrad equation to evaluate the test rather than the function (8, p. 65). Davey, supporting Spearman's theory says, "A verbal mental test measures the same general factor '*g*' as does a test similar in form but non-verbal in material" (1, p. 47). Dearborn, without attempting to lend support to any underlying presupposition, says, "One cannot differentiate between grades of intelligence on the basis of materials employed, as, e.g., between the handling of things or ideas, whether the latter are said to be concrete and abstract, or between the handling of things and the symbol of things, such as words or numbers" (2, p. 12).

Which of the two alternative interpretations is correct must depend for its verdict upon the results of experience. Before judgment can be pronounced the first essential is to give the subject representative tests of equal difficulty from both fields; the next essential is proof that the same function is brought into play in facing the two kinds of tools, ideas, and things. A blanket term like "linguistic ability" has no significance in itself unless it resolves itself into some form of completion test, analogy test, comprehension test, or some such; and even then it is useless unless the availability of the analogies or situations in the child's previous experiences is considered. So with non-linguistic ability; the phrase is nothing more than a generalization for a number of abilities each of which should be subjected to separate measurement. Until these are separately treated and the function each formboard measures is known, the totality of these measures, so-called formboard ability, is of very little significance.

Formboards may be divided into two groups:

1. Those which measure one's ability to judge the appropriateness of parts to parts, or of parts to wholes.
2. Those which analyze wholes into parts whose relations to one another are perceived.

Typical of the first type are the Gwyn Triangle Test, the Two and Five Figure Formboard Tests, Healy A, and the Worcester Series. Ex-

amples of the second type are the Ship Test, the Manikin Test, the Drugstore Test, Healy P.C. I and II, and Worksample No. 5 of Johnson O'Connor's Mechanical Aptitude Test (6, p. 29). It is very true that the tests measure speed and accuracy of muscular movement as well as adaptation to novel situations, but these are incidental results, secured from observation of the subject's method of approach. The factor at stake in the evaluation of one's ability is the strength of the function being tested.

There seem to be indications that the same function may operate in both linguistic and non-linguistic tests, and that the discrepancy in test results, when the two kinds are given the same subject, are due to the tests not having common elements, and the occasional possibility of the same function being stronger in some people than in others. There does not seem to be sufficient proof for Spearman's hypothesis that the two types of tests measure a common function. Each type of test measures a different function, and when that function is at work, be it in the form of linguistic or non-linguistic ability, its manifestation is the same, provided the tests are of equal value.

Support of this thesis comes from an analysis of varied examinations of the two types given the same person. This has been secured in 25 cases. The Dearborn Group Test, Series II, was given a group of 300 students in a high school, and then on the basis of their showing marked strength or weakness in parts involving speed or the perceptions of relations, 25 subjects were selected for further examination. The distinction between linguistic and non-linguistic was not observed; the point at stake was the contrast between speed in recall and ability to see relations. Tests 5 and 6 of Dearborn, Series II, were regarded as speed tests, and Test 3 one of relations. The 25 subjects were given the following additional tests: Stanford-Binet, Kohs Block Design, and the Macquarrie Test for Mechanical Ability. The Binet Test was chosen because of its being predominantly linguistic; the Kohs (3) was chosen because of its claim to measure ability to synthesize, and because of a striking correlation between failure on that test and on tests which measure ability to perceive relations. Preliminary trials seemed to indicate that it measures a function similar to that at work when solving Dearborn, Test 3. The Macquarrie Test was chosen because of its testing two things which are regarded as essential in mechanical work, speed and accuracy of movement, and appreciation of exact relationships (4, pp. 15-24). To measure the first, Macquarrie has four tests, Nos. 1, 2, 3, and 7; the remaining three tests measure the other function. These two divisions of the test are not indicated by the author, but preliminary trials showed such a marked discrepancy in the two groups that it was thought advisable to see with what other types of tests these results correlated. The results from the 25 subjects who took the Macquarrie tests are shown in Table 1.

TABLE 1

Case	Age	Raw score Speed and accuracy tests					Raw score Relation tests			
		1	2	3	7	Av.	4	5	6	Av.
1 (E.P.)	14	28	39	22	12	25	13	1	5	6
2 (R.S.)	14	30	56	64	44	48	30	40	50	40
3 (E.A.)	16	45	60	81	46	58	30	50	32	37
4 (E.G.)	15	32	40	72	33	44	23	36	18	26
5 (H.N.)	16	45	56	70	73	61	25	30	30	27
6 (H.A.)	14	45	50	70	55	54	60	80	66	69
7 (C.N.)	13	33	40	80	26	45	33	50	0	28
8 (G.W.)	13	30	33	52	20	34	18	28	33	26
9 (E.B.)	14	45	52	70	50	50	65	75	67	69
10 (C.O.)	18	45	50	85	55	58	27	33	20	29
11 (C.H.)	15	40	50	80	48	54	42	20	25	29
12 (S.J.)	15	65	75	85	54	70	30	50	10	32
13 (R.M.)	14	30	50	55	20	39	60	40	20	50
14 (W.B.)	13	25	35	50	35	36	25	50	20	24
15 (F.L.)	14	33	50	75	50	54	60	50	20	32
16 (E.B.)	14	30	60	80	75	61	30	30	10	23
17 (R.L.)	14	25	50	60	25	40	25	50	10	21
18 (J.F.)	14	50	65	75	50	60	40	35	18	31
19 (E.M.)	17	30	60	70	40	52	20	22	18	20
20 (M.W.)	16	40	60	75	54	60	55	55	33	41
21 (U.C.)	16	50	40	60	40	47	35	50	10	22
22 (G.M.)	14	40	45	55	50	47	15	50	33	25
23 (R.H.)	15	50	50	90	60	62	35	55	35	41
24 (F.W.)	14	25	33	45	20	30	20	50	25	33
25 (L.B.)	14	50	45	75	50	55	65	75	50	63

Inspection of the averages leads to a classification of the subjects in two groups:

1. Those whose average on speed and accuracy tests is above that on the relation tests.

2. Those whose average on the speed and accuracy tests is equal to or less than that on the relation tests. In Tables 3 and 4 the averages for these two groups on the Macquarrie test are given under the headings of Speed and Relation. In addition to these the scores (in percentage terms) of the same subjects on Dearborn 1, 3, and 4, Kohs Block Design, and selected tests from the Stanford-Binet are given. Dearborn 1 and 4 are averaged and compared with Dearborn 3. In Tests 1 and 4 speed of recall is the most important factor; in Test 3 an apprehension of relations may be necessary. The Stanford-Binet tests of Linguistic and Mechanical Ability with which these are compared are taken from Riley's results (7). He gave the Stanford-Binet, the Arthur Performance Scale, and the Minnesota Assembly Test to 65 boys from ages 10 to 17 years, and, on the basis of high correlations between subjects on the S-B and high or low

TABLE 2

Year	Mechanical Group	Linguistic group
10	Test 1 (Vocabulary) Test 3 (Designs)	Test 5 (Comprehension) Test 6 (60 words)
12	Test 2 (Ball and field) Test 5 (Fables) Test 7 (Pictures)	Test 4 (Dissected sentences) Test 8 (Similarities)
14	Test 2 (Induction) Test 5 (Arithmetic)	Test 4 (Problems)
16	Test 4 (Problem boxes) Test 6 (Code)	Test 1 (Vocabulary) Test 5 (Digits backwards)
18	Test 2 (Paper cutting)	Test 3 (Digits backwards)

TABLE 3

Case	Averages for			Stanford-Binet			Kohs Block Design Patterns compl.
	Speed	Relation	Dear- born 1 and 4	Dear- born 3	Mechan.	Linguis.	
1	25	6	48	14	10-3 12-3, 7	all	4
2	48	40	42	60	16-4, 6	14-4	12
3	58	37	67	33	14-2 16-4	16-1, 5	9
4	44	26	60	52	16-4 16-6	16-1 16-5	10
5	61	27	52	60	16-6	16-1, 5	13
7	45	28	45	33	12-3, 5	14-4	8
8	34	26	42	0	10-3 12-3, 5, 7	12-4, 8	3
10	58	29	67	50	16-6 18-2	16-5 18-3	10
11	54	29	46	40	14-2 16-4, 6	16-1	11
12	70	32	62	15	12-5 14-2, 5	14-4 16-1, 5	5
14	36	24	50	20	12-5, 6 14-2, 5	14-4	6
15	54	32	46	55	16-6	14-4 16-1, 5	12
16	61	23	50	33	12-5, 7	16-1, 5	8
17	40	21	50	30	12-5, 7	14-4 16-1, 5	8
18	60	31	36	0	10-3 12-3, 5, 7	14-4	7
19	52	20	60	30	14-2 16-4, 6	16-1, 5	8
20	60	41	50	25	14-2	16-1, 5	8
21	47	22	38	25	14-2	16-1, 5	9
22	47	25	57	30	14-2	16-5	10
23	62	41	27	30	12-5 14-2	14-4	9

scores on the mechanical tests, he selected as the best indicators of mechanical and linguistic ability the parts of the S-B test shown in Table 2.

In Tables 3 and 4 the records on the S-B test have been limited to first failures on the type of ability indicated. Where tests from any year are mentioned in either column, and the other mechanical or linguistic indicators are omitted, the implication is that the remaining indicators in that year have been completed, but that none have been done beyond that one.

In Table 4, according to the same plan as Table 3, the results are given for those cases where the average scores for speed were equal to or less than the score on Test 3 of the Dearborn test.

TABLE 4

Averages for					Stanford-Binet		Kohs Block Design Patterns Patterns compl.
Case	Speed	Relation	Dear- born 1 and 4	Dear- born 3	Mechan.	Linguis.	
6	54	69	40	75	14-2 16-4	14-4	14
9	50	69	36	50	12-5 14-2, 5	12-8 14-4	9
13	39	50	42	25	14-2	14-4	7
24	30	33	30	25	12-5	12-8	8
25	55	63	70	100	16-6 18-2	16-5	16

In the Kohs Block Design Test there are 17 patterns of increasing difficulty to be completed, The results for the 25 subjects are given in Tables 3 and 4.

Summarizing the results from these two tables, we find:

1. In the 20 cases in which speed exceeded relation on the Macquarrie Test, 17 scored a higher average on Dearborn 1 and 4 than on Dearborn 3, whereas in the 5 cases in which relation exceeded speed, 60% made a much higher score on Dearborn 3 than on Dearborn 1 and 2.

2. In the so-called mechanical group, Table 3, the mechanical indicators on the Stanford-Binet showed 3 to be advanced 2 years, 2 advanced 1 year, 1 the same mechanical as chronological age, 2 retarded 1 year, 8 retarded 2 years, 2 retarded 3 years, and 2 retarded 4 years; whereas in the linguistic group, the mechanical indicators showed 2 to be retarded 2 years, 1 the same mechanically as chronologically, 1 advanced 1 year, and 1 advanced 2 years.

3. The linguistic indicators showed 2 to be advanced 2 years, 4 advanced 1 year, 7 the same linguistic as chronological age, 5 to be retarded

1 year, and 2 to be retarded 2 years in the mechanical group; whereas 1 was advanced 2 years, 1 one year, 2 the same, and 1 retarded 2 years in the linguistic group.

4. The Kohs test results show a significant correlation with a low or high score on Dearborn 3. The four lowest scores on Dearborn 3, cases 1, 8, 12, and 14, were the lowest on the Block Design Test. None of them completed more than 6 patterns. On the other hand, the highest scores in Dearborn 3 were among the highest scores in Kohs. Cases 2, 5, 6, 19, 11, 15, and 25, the seven highest in Dearborn 3, all completed 10 or more patterns in Kohs.

The results from these four types of tests seem to indicate that the same function is at play in similar aspects of all four. If a person makes a high score on Dearborn 3, he will likely make a higher score on Macquarrie 4, 5, and 6 than he will on 1, 2, 3, and 7; a higher score in the linguistic parts of the Stanford-Binet than in the mechanical parts, and a fairly high score on Kohs. If he makes a low score in Dearborn 3, it will likely be bettered in Dearborn 1 and 4, Macquarrie 1, 2, 3, and 7, and by the Stanford mechanical indicators. A high score in these, however, is not a necessary accompaniment of a low score in Dearborn 3; there are as many cases where the score is on a par as there are significant differences.

The cases are too few to warrant a definite conclusion, but the evidence points to the presence of two functions: first, a function, common to related parts in all four tests, which is responsible for speed of recall and speed in muscular movements; and, secondly, a function which appears in the combining of separate parts according to a new pattern, dividing up wholes according to a plan, or seeing the relation of one part to the remaining parts in a whole. Both functions seem to operate in both linguistic and non-linguistic ability; our results indicate that strength in one is not necessarily correlated with strength in the other.

REFERENCES

1. DAVEY, F. A comparison of group verbal and pictorial tests of intelligence. *Brit. J. Psychol.*, 1926, **17**, 27-48.
2. DEARBORN, W. F. A series of form board tests of intelligence. *Harvard Monog. Educ.*, 1923, **1**, No. 4.
3. KOHS, S. C. Intelligence measurement. New York: Macmillan, 1923. Pp. xii+312.
4. MACQUARRIE, T. W. Test for mechanical ability. *J. Person. Res.*, 1927-1928, **6**, 15-24.
5. MATEER, F. The unstable child. New York: Appleton, 1924. Pp. xii+471.
6. O'CONNOR, J. Born that way. Baltimore, Md.: Williams & Wilkins, 1927. Pp. 296.
7. RILEY, G. Stanford-Binet indicators of mechanical ability. *Psychol. Clin.*, 1929, **18**, 128-132.

8. SPEARMAN, C. The abilities of man. New York: Macmillan, 1927. Pp. xxiii+415.
9. THORNDIKE, E. L. Symposium on intelligence and its measurement. *J. Educ. Psychol.*, 1921, **12**, 124-127.
10. WELLS, F. L. Mental tests in clinical practice. Yonkers, N. Y.: World Book Co., 1927. Pp. x+315.

Mount Holyoke College
South Hadley, Massachusetts

A QUANTITATIVE INVESTIGATION OF EARLY MEMORIES

MASON N. CROOK AND LUBERTA HARDEN

This investigation was designed to test the hypothesis that repression of childhood memories is associated with neurotic or unstable personality traits. If the neurotic individual does tend to repress early memories, then the total number of memories from any given period of childhood should be correlated with freedom from neurotic traits. In this experiment data on the number of memories and on neurotic tendencies in a group of students were examined for evidence of such association. Data on ability in immediate recall of nonsense material, which were available on the same individuals, were also correlated with the data on ability to recall from childhood.

The data on ability to recall nonsense material were obtained in the course of an experiment on retroactive inhibition. Lists of five pairs of nonsense syllables were exposed five times (.5 second per syllable), and immediate recall asked for. The right associates method was used, and the first syllables were supplied, the subject being required to write the second syllables in the appropriate places. After a two-minute interpolation period spent either in reading ordinary prose without disturbance (control situation) or in reading the same prose with electric shocks, the firing of a revolver, and the sudden loss of support produced by a falling chair-seat (test situation), recall was again called for.¹

The records used in this study are (a) the difference between the time required to recall one syllable after interpolation with disturbance and the time required to recall one syllable after interpolation without disturbance (designated "time difference" in Table 1); (b) percentage of syllables correctly recalled after interpolation without disturbance (delayed recall); and (c) percentage of syllables correctly recalled immediately after study (immediate recall).

The Pressey X-O test was given to the same individuals who served in

¹For a description of this procedure, see (1).

the retroactive inhibition experiment, and the total affectivity scores were taken as the measure of neuroticism.

Data on early memories were obtained by requesting each subject to fill out a mimeographed chart bearing the following instructions.

"The accompanying chart is to be filled out by placing on it a cross for each memory you have from the first six years of your childhood. The chart is scaled in two dimensions; the vertical scale represents time and the horizontal scale represents certainty. If you remember an incident which you think occurred at 3 years and 9 months, place a cross opposite the 3-year-9-month level on the age scale, and under the number on the horizontal scale which represents your degree of certainty that you have judged the time correctly within one year, that is, indicate your degree of certainty that the event occurred between three years three months and four years three months.

"The fifty point on the certainty scale for any age level represents an even chance that an event occurred at that age. The hundred point represents absolute certainty that the event occurred at that age, and the intermediate points represent intermediate degrees of certainty.

"The chart is scaled in large units. Approximate intermediate points as closely as you can.

"Put only one cross for each incident, but remember as many incidents as you can.

"Before recording each incident, check the time when it occurred by any secondary information you may have. The dates of visits, movements from one house to another, births of brothers and sisters, etc., are useful for this purpose.

"For each incident of which you can get verification from records or from other people, place a circle around the cross.

"If you have memories whose genuineness you are uncertain of, state their number, the supposed times of their occurrence, and the reason for your uncertainty on the reverse of this sheet. Use extra space for additional remarks, if any.

"Please sign your name."

From the results on the early memory charts the following scores were computed for each subject: (*a*) total number of memories from the first six years age of life, (*b*) age of subject at time of earliest memory, (*c*) age of subject at time of earliest memory of 75-100 certainty, and (*d*) age of subject at time of earliest memory of 90-100 certainty.

Total number of memories ranged from 0 to 140; the 140 was an extreme case, the second highest being 40. The subject making the score of 140 will be designated as G in the following discussion. Certain of the correlations were computed with G's score included and recomputed with this score excluded.

A few of the subjects indicated a number of specific memories and stated

that they had others which there was not room on the chart to indicate specifically. The scores allotted to these subjects for the purpose of working the correlations were in each case the lowest that the statement of the subject might be interpreted to designate.

The earliest memories ranged from nine months to over six years. Each subject having no memory within the six-year period was allotted a score of six years for the purpose of working the correlations.

The correlations computed are listed in Table 1. For those correlations having the highest values the critical ratios and corresponding chances in 100 are also given.

The correlations between early memories and recall of nonsense syllables are insignificant in magnitude. The reliability of the scores on immediate recall, however, is only .47, as indicated by the intercorrelation of the two series, so negative results here do not show with certainty that no relationship exists.

The difference between time for recall of first syllable with disturbance and time for recall of first syllable without disturbance (designated as

TABLE 1

Items correlated	<i>N</i>	<i>r</i>	Critical ratio	Chances in 100
Pressey scores and total memories (G excluded)	18	-.354	1.7	96
Pressey scores and total memories (G included)	19	-.37	1.7	96
Pressey scores and earliest memories	19	.52	2.65	99
Pressey scores and earliest memories (75-100)	19	.32	1.55	93
Pressey scores and earliest memories (90-100)	19	-.17	.79	78
Time difference and total memories	28	-.17	.94	83
Time difference and earliest memories	29	-.07		
Time difference and Pressey scores	18	.08		
Delayed recall and total memories (G excluded)	30	.16		
Delayed recall and earliest memories	31	.12		
Delayed recall and earliest memories (75-100)	31	.14		
Immediate recall and total memories	30	-.08		
Immediate recall and total memories (75-100)	31	-.02		
Immediate recall (1) and immediate recall (2)	32	.47		

"time difference" in Table 1), which might be considered an index of emotionality or disturbability, correlates to an insignificant degree with the Pressey scores and also with early memories.

The one group of correlations which shows values of significant magnitude is that between Pressey scores and the various early memory scores. The Pressey scores correlate $-.35$ with the total number of memories (G excluded; with G included the correlation is raised to only $-.37$). There are 96 chances in 100 that the true correlation here is greater than 0 in the negative direction. The correlation obtained between Pressey scores and age at time of earliest memory is $.52$, with 99 chances in 100 that the true correlation is greater than zero. Therefore the tendency is as follows: The more emotionally stable an individual is, as indicated by a low Pressey score, the greater number of memories he retains from the first six years of his life and the earlier the age from which he retains the first memory. However, as the Pressey scores are correlated against earliest memories of higher degrees of certainty in time location, the size of the correlations decreases. The correlation with earliest memories of 75-100 certainty is $.32$ (chances 93 in 100); the correlation with earliest memories of 90-100 certainty is $-.17$ (chances 78 in 100).

In summary, we find that neuroticism as measured by Pressey scores correlates negatively with wealth of early memories as reported by the subjects in this investigation. The correlations found, while not highly reliable, are at least suggestive of a genuine relationship. We also find that the correlation between Pressey scores and age at time of earliest memory decreases as the subjects' certainty of the time placement increases. These facts considered together are capable of two possible interpretations: (a) that the relationship between neuroticism and survival of memories is effective only with respect to those memories which are vague and uncertain for other reasons, or (b) that the vagueness of certain types of memory is itself due in part to neurotic tendencies, complete repression being the result of the same tendencies when present in more pronounced degree.

Previous investigations have been directed at the relative permanency of pleasant and unpleasant memories, and the results of different studies are in apparent conflict. It is suggested that more work along the lines here reported might be fruitful, because of the great importance attached to the early years in psychoanalytic theory.

REFERENCES

1. HARDEN, L. M. The effect of emotional reactions upon retention. *J. Gen. Psychol.*, 1930, **3**, 197-221.
2. MELTZER, H. The present status of experimental studies on the relationship of feeling to memory. *Psychol. Rev.*, 1930, **37**, 124-139.

Dartmouth College
Hanover, New Hampshire

Clark University
Worcester, Massachusetts

BOOKS

HAROLD D. LASSWELL. *Psychopathology and Politics*. Chicago: Univ. Chicago Press, 1930. Pp. ix+285. \$3.00.

Dr. Lasswell's newest venture in the application of psychoanalytic findings and techniques to sociology is built around four chapters (VI-IX) of case histories, selected from a much larger mass of material collected, presumably by him, chiefly from essentially normal people, many of them occupying positions of political significance. To these he has prefixed orienting chapters introducing the unsophisticated reader to the most relevant findings of the general method, and suffixed others, some of them expanded from material published elsewhere, on methodology (chiefly objectification), and on a theory of the state. A good deal of historical material, usefully summarizing German theory on politics and characterology, is interspersed.

A writer on these topics is handicapped by a troublesome dichotomy in his audience: the reader is likely to be either totally unsophisticated, in which case the whole business seems to him far-fetched and fantastic, or acquainted with the nature and functioning of unconscious motivation, in which case he is likely to be impatient at the laboring of the obvious. It is difficult to evaluate the book for this reason—at least its main portion. The reviewer happens to fall in the latter class, and so it is as difficult for him to maintain interest during the discussion as it would be, say, during a lengthy proof that at a sufficiently high temperature teak and mahogany, as well as the more common oak and maple, combine with oxygen; but he is aware, of course, that there are myriads of estimable and intelligent people who cannot be convinced that the entire body of their motives is not transparent to them.

It is somewhat different with the methodological portions, although, as with much sociology—particularly German—the theory has a constant tendency to wander off from the facts and get lost in the fascination of its own self-consistency. I found, for example, the concept of tension level in the chapter "The Politics of Prevention" a valuable addition to my thinking; but I could not understand what most of the rest of the chapter was doing around it. Is not the essential chapter contained in a half-dozen sentences like these?

Our problem is to be ruled by the truth about the conditions of harmonious human relations, and the discovery of the truth is an object of specialized research; it is no monopoly of people as people, or of the ruler as ruler The politics of prevention does not depend upon a series of changes in the organization of government. It depends upon a reorientation in the minds of those who think about society around the central problems: What are the principal factors which modify the tension level of the community? What is the specific relevance of a

proposed line of action to the temporary and permanent modification of the tension level?

In somewhat similar fashion, the chapter on "The Prolonged Interview and Its Objectification" boils down to the necessity for the collection of objective records of just what happens in the psychoanalytic hour, in order that correlations may be unearthed between parts of the subject's behavior and between the latter and the behavior of the analyst; and we have known that for a long time—the trick is to find the resources. It is a pleasure to observe from Dr. Lasswell's footnotes, etc., that work is evidently under way in this field at Chicago under the direction of Dr. Harry Stack Sullivan, one of the country's most astute psychiatrists.

The chapter on "The Personality System and its Substitutive Reactions" contains the suggestion, highly meritorious but familiar from the work of Lewin, that on a tendency hypothesis much should be ascertainable by opposing standard checks to the principal drives and observing the types of substitutions. There is also a suggested rating scale for political behavior, and a comment to the effect that until overt symbolisms are correlated with careful intensive studies of their meaning on numerous individuals, the value of rating studies by raters of relatively slight acquaintance with their subjects is unknown and probably small.

There is a selected and annotated bibliography and a list of questions designed to elicit the political experience of the reader.

RAYMOND R. WILLOUGHBY

Clark University
Worcester, Massachusetts

CLAPARÈDE, E. *L'éducation fonctionnelle.* (Collection d'Actualités Pédagogiques.) Neuchâtel & Paris: Delachaux & Niestlé, 1931. Pp. 266. Fr. 4.50.

FISHER, V. E., & HANNA, J. V. *The dissatisfied worker.* New York: Macmillan, 1931. Pp. x+260. \$2.00.

FOLSOM, J. K. *Social psychology.* New York: Harper, 1931. Pp. xviii+701. \$3.50.

HIRSCH, N. D. M. *Genius and creative intelligence.* Cambridge, Mass.: Sci-Art, 1931. Pp. 339. \$4.50.

KAHN, M. C. *Djuka: the Bush negroes of Dutch Guiana.* New York: Viking, 1931. Pp. xiv+234. \$3.50.

KATZ, D., & ALLPORT, F. H. *Students' attitudes: a report of the Syracuse University reaction study.* Syracuse, N. Y.: Craftsman Press, 1931. Pp. xxviii+408. \$3.50.

MATHEWSON, S. B., LEISERSON, W. M., DENNISON, H. S., & MORGAN, A. E. *Restriction of output among unorganized workers.* New York: Viking Press, 1931. Pp. xii+212. \$3.00.

MEAD, M. *Social organization.* (Bernice P. Bishop Museum, Bull., 76.) Honolulu, Hawaii: Bernice P. Bishop Museum, 1930. Pp. ii+218.

- MURPHY, G., & MURPHY, L. B. Experimental social psychology. New York & London: Harper, 1931. Pp. xii+709. \$3.50.
- POULL, L. E., BRISTOL, A. S., KING, H. B., & PEATMAN, L. B. The Randall's Island performance series: two, three, and four year tests. New York: Columbia Univ. Press, 1931. Pp. vi+26. \$1.00.
- REED, H. B. The influence of training on changes in variability in achievement. *Psychol. Monog.*, 1931, 41, No. 2 (Whole No. 185.) Pp. 59.
- REUTER, E. B. Race mixture. New York: Whittlesey House (McGraw-Hill), 1931. Pp. viii+224. \$2.50.
- ROBACK, A. A. Personality: the crux of social intercourse. Cambridge, Mass.: Sci-Art, 1931. Pp. 144.
- SCHEEERER, M. Der Lehre von der Gestalt: ihre Methode und ihr psychologischer Gegenstand. Berlin & Leipzig: Gruyter, 1931. Pp. iv+405.
- SELIGMAN, E. R. A., & JOHNSON, A. [Editors.] Encyclopaedia of the social sciences. Volume III: Bri-Com. New York: Macmillan, 1930. Pp. xxii+681. \$7.50.
- THOMAS, N. America's way out: a program for democracy. New York: Macmillan, 1931. Pp. xii+324. \$2.50.
- THURSTONE, L. L., & JENKINS, R. L. Order of birth, parent-age, and intelligence. (*Behav. Res. Fund. Monog.*) Chicago: Univ. Chicago Press, 1931. Pp. xiv+135. \$3.00.
- VAISSIÈRE, J. DE LA, SERTILLANGES, A. D., FARGUES, M., JAOUEN, J., FAUVILLE, A., & BUYSE, R. Questions actuelles de pédagogie. Vol. 1. Seine-et-Oise: Juvisy, Éditions du Cerf, 1931. Pp. 198. 10 fr.
- WARDEN, C. J., JENKINS, J. N., WARNER, L. H., JENKINS, M., HAMILTON, E. L., & NISSEN, H. W. Animal motivation: experimental studies on the albino rat. New York: Columbia Univ. Press, 1931. Pp. xii+502.
- WOODWORTH, R. S. Contemporary schools of psychology. (*Psychology Series*, Ed. by A. T. Poffenberger.) New York: Ronald Press, 1931. Pp. vi+232. \$2.50.
- YOUNG, K. [Ed.] Social attitudes. (*Amer. Soc. Science Ser.*) New York: Holt, 1931. Pp. xii+375. \$3.00.

THE PSYCHOLOGY OF INCENTIVES*

From the Department of University Extension, Columbia University

ELIZABETH B. HURLOCK

In recent years, psychologists have given much attention to experimental studies of how human beings and animals learn. Learning curves, showing the rate and method of learning, are constantly appearing in psychological literature, so that today knowledge of how learning takes place is an important phase of the field of psychology. Accompanying these studies have been analyses of the capacity for learning in both human and animal subjects. Intelligence tests, and tests for different aspects of mental ability, have been used to show why it is that one subject learns more quickly and economically than another. But, back of all of this is the will to learn, or the desire on the part of the subject to master the task set before him. Only very recently has the psychologist turned his attention to the study of this very fundamental and important aspect of learning. Recognition is fast being given to the fact that learning is not dependent alone on ability to learn, or on correct methods of learning, but that some motive or incentive is essential if the learner is to master the task assigned him.

The term "incentive," as used by the psychologist, refers to any *externally* controlled condition which stimulates activity on the part of a human being or animal. It is distinguished from the term "motive" which relates primarily to those forces from within the individual which impel action on his part. While it is true that this distinction is to a certain extent an arbitrary one, due to the fact that it is impossible at all times to differentiate between an incentive and a motive, since in many instances motives originate from incentives, nevertheless, an attempt has been made in this study to summarize only those experimental works which attempt to measure the influence of incentives on human or animal behavior.

RELATIONSHIP OF ABILITY TO ACHIEVEMENT

Since the intelligence test has come into common use in schools

*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office, October 1, 1930.

and colleges as a means of measuring the mental ability of children, the psychologist has discovered that there is in the case of many students a wide discrepancy between the mental capacity of the child, as measured by the intelligence test, and his actual achievement, as measured by school grades and achievement tests. This discrepancy is found in most cases to be greatest in children of the higher levels of intelligence. The reason that their achievement does not measure up to their abilities is almost always ascribed to lack of necessary incentive on their part.

Many studies have been made on school children and college students to determine how great the discrepancy is between ability and achievement. Torgerson and Sherman (126) gave 216 children with MA's of 12, and IQ's ranging from 72 to 136, intelligence tests and Thorndike-McCall Reading Tests to determine their accomplishment quotients. The results of the experiment showed that the children in the backward group (76-85 IQ) read as well as children of normal intelligence (96-105 IQ), and that both of these groups were above the superior pupils (116 IQ and above). Hornell Hart (60) showed that if children were allowed to progress in school at the rate justified by their ability, they should graduate on the average from high school at the age of 15 years and 5 months instead of 17 years and 3 months.

Poffenberger and Carpenter (101), starting with the assumption that school achievement involves factors other than those measured by intelligence tests, gave the Carnegie Revision of the Downey Will-Temperament Test to 97 children of the sixth, seventh, and eighth grades. They found that the character traits contributing to success in school work are swift coordination and decision, motor impulsion and flexibility, speed of movement and freedom from load, care for detail and lack of motor inhibition, and flexibility and high assurance. On the other hand, those traits which contribute to failure are slow coordination, lack of motor impulsion and flexibility, low speed of movement and freedom from load, lack of care for detail and motor inhibition, flexibility and low assurance.

Pintner and Marshall (99) made the following statement, based on evidence obtained from the Pintner Mental-Educational Survey Tests:

"As a matter of fact, we shall see what we have long suspected and what we have only recently consciously realized, that the greatest amount of waste exists among the brighter

pupils of a class or among the better schools in a school system. It is, as a rule, the more intelligent pupils who are working below capacity, even when they are keeping well up to the average of the class. We have been pushing and cramming the duller children, while the brighter ones have been allowed to loaf. The bright child is the most retarded in our schools. The dull child is the most accelerated."

Toops and Symonds (125) comment on this statement thus:

"Is it not a remarkable coincidence that the 'below' normal in intelligence are for the most part above in motivation, while the 'above' normal in intelligence are for the most part below average in motivation?"

McPhail (90) and Richards (106) found discrepancies in ability and achievement in school children similar to those of the above writers. In an attempt to predict high-school success, Rector (104) studied 431 tenth-grade children. He found a correlation of .28 between intelligence test scores and combined marks for all school subjects.

Similar studies have been made on college students. McCrory (92) found correlations of .465 and .457 between the scores made on Otis Self-Administering Tests and high-school and college grades, respectively. Rosenow (107) found a correlation of .44 between intelligence test scores and scholarship records for 883 University of Kansas students at the end of their freshman year, while May (90) found a correlation of .60 between intelligence and "credit points" in the case of 450 freshmen at Syracuse University. Pyle (102) found a correlation of .52 between intelligence test scores and academic success for 140 college students.

In industry, there are indications that the workers are not working up to their maximum abilities, as is true of college students and school children. Poffenberger (100) has discussed this problem in a general way, while Kitson (80) has shown through a study of 40 experienced hand compositors that when some incentive to improve their work was offered, in all except one case, the individual output of the workers was increased.

ANIMAL INCENTIVES

Experimental studies of animal learning have given much attention to the question of what incentives are the most effectual in bringing about action in animals when placed in a given learning

situation. Food, electric shock, and sex are the three which have most frequently been used by animal psychologists. In many instances, there has also been an attempt to evaluate the various incentives employed to see which is the most effective.

Washburn (136) studied the rate of learning of 20 mice in a maze situation when the only incentive used was milk. A correlation of $-.41$ between hunger and speed of running led her to conclude that "not the most important result of this study is the emphasis with which it shows that animals should be studied individually." Dashiell (31) likewise measured the strength of the hunger stimulus in white rats in running mazes. He found that the average number of blocks in the maze covered by fed rats was 26.7 as opposed to 42.9 covered by the "hungry" group. Blodgett (11), in an attempt to study the efficiency of practice when unaccompanied by reward, divided his rats into two groups, the first of which received reward of food only during the latter part of the experiment, while the second or control group was rewarded throughout the entire experiment. The rats learned more slowly when unrewarded, and, in the experimental group, a great improvement was noted when the reward was suddenly introduced.

To test the amount of reward necessary to learning in chicks, Grindley (55) divided 50 chicks into five groups of ten each. The problem set for them to learn was running out of a release box to a tray at the end of a passage remote from the box, and from which they could obtain food. The rewards given the chicks were as follows: group one, 1 grain of boiled rice; group two, 2 grains; group four, 4 grains; group six, 6 grains; and group 0, no reward at all. It was found that "most of the chicks of groups four and six walked or ran briskly down the passage, showing some signs of excitement, while the chicks of groups one and two usually proceeded in a more leisurely way."

The influence of the length of the hunger period on learning has been extensively studied. Watson (138) found that no significant difference in learning took place when rats, accustomed to working in a maze by the delayed feeding method were fed immediately after each trial, nor were those accustomed to immediate feeding influenced by delayed feeding. Haas (56), Warden and Haas (131), and Hamilton (57) followed up Watson's (138) method. Delay periods of 15 seconds to 5 minutes were used. The 5-minute interval was found to have "no measurable effect upon the rate of fixation,"

while the hunger drive was "markedly decreased by a period of delay as short as 15 seconds." Moss (96) measured the strength of hunger drives in rats varying from 12 to 148 hours. He found that in 8 out of 10 cases the hunger drive was strong enough to overcome resistances in learning, and that rewards were more effective than punishments. Szymanski (120, 121), using the maze problem to study the effect of different hunger periods on learning, found that the most effective learning occurred with the 24-hour period, while the least effective one was the 30-minute period.

Holden (64) studied the effect of different degrees of starvation on learning when some obstruction (in her experiment, an electric grill which the rats had to cross) was introduced. The results of her experiment showed that "the hunger drive appears to increase from 12 to 36 hours and to decrease with further increases in the length of the starvation period." Williams (146) found that a conditioned food stimulus acted as an incentive to learning for a time, but that its incentive value decreased soon unless it was reinforced. Sharp (111) discovered that in the case of rats "more or less continuous running of the maze with food incentive present introduces appreciable disintegrative effects. Removal of the food incentives introduces a marked degree of disintegration into a maze habit, which practice within the limits of our experiment reduces but little."

Thirst as an animal incentive has been studied by Warner (134). Using the obstruction method described by Jenkins, Warner, and Warden (73), he found that, in the case of white rats, after one day of water deprivation the rats showed a greater tendency to cross the electrical obstruction, while after that time, the tendency decreased constantly until the death of the animal.

Glaser (53) studied the influence of water as an incentive to maze learning for white rats. A tank was filled with warm or cold water, to which salt was sometimes added. When the rats were suddenly thrust in the tank, they made strenuous efforts to escape. It was found that, in 83% of the cases, the time taken to escape was decreased when this incentive was used.

The measurement of the sex drive as an animal incentive through the use of the obstruction method has been made by Jenkins (72) and Warner (133). Jenkins found that one day of segregation doubled the number of crossings of the electrical obstruction on the part of the male, and that the relative effect of sex segregation is

greater in males than in females. Warner's problem closely resembled that of Jenkins. In measuring the relative effectiveness of the sex drive in male and female rats, he found that the female "in oestrus displays more activity in the form of crossing the electrical obstruction to the incentive compartment even though that compartment be empty than does the female in dioestrus even though the compartment contains a male. A male rat which has not mated for twenty-four hours will cross electrical obstruction less often to an empty incentive compartment than will a male of any of the sex deprivation intervals studied (0, 6, 12 hours, 4, 7, 28 days) to an incentive compartment containing a female in oestrus."

Many of the studies of animal incentives have had as their main purpose the measurement of the relative effectiveness of certain commonly used incentives in animal learning. The relative value of reward, usually in the form of food, and punishment, in the form of electric shock, has received the greatest amount of attention. Hoge and Stocking (63), using the visual discrimination problem for rats, concluded that it "seems evident from the experiment that a combination of reward and punishment is more effective than either alone. Punishment is more effective than reward so far as the rate of learning is concerned." Aylesworth (5 and 130), in a study similar to that of Hoge and Stocking (63), found likewise that reward and punishment together are better than either alone, and that, of the two, punishment is more effective than reward.

Reward and punishment were used by Casteel (22) in a study of the discriminative ability of the painted turtle. The electric shock as punishment had different effects on the different turtles. Some were made wild and nervous, others became balky and refused to do anything, while others became so accustomed to the electric shock that its incentive value disappeared. Dodson (37), in a visual discrimination problem with rats, found that "the electric shock is more favorable to the learning process in the white rat than is hunger in the case of a simple discrimination problem."

Theoretical discussions of the relative values of reward and punishment in animal learning, based on experimental studies, have been given by Watson (137), Washburn (135), and Hunter (67, pp. 17-39). Washburn concludes that "punishment appears to produce more rapid learning than reward, unless it is so severe that it attaches itself to the whole learning situation. Punishment and reward combined give, probably, better results than either alone."

Watson lays stress on the fact that animals differ greatly in their sensibility to punishment.

The relative values of food and sex as incentives have been studied by several experimenters. Chiao Tsai (129) found that in the case of rats, food was chosen 77% of the time as an incentive in maze learning, while the sex object was chosen only 23% of the time. Warner (132) found that the average number of crossings of the electrical obstruction placed in the maze was 13.45 to the sex object and 19.1 to food. His conclusion was that "we feel justified, then, in saying that, using the obstruction method, the hunger drive is stronger than the sex drive in the male animal." In the case of the female rat, the difference was also in favor of the hunger drive, but this difference was not large enough to be statistically reliable.

Stone and Sturman (117) did not find a difference in favor of hunger as an incentive. From their experiment on maze learning in rats, when the food and sex incentives were used, they concluded that "it would seem that one is justified in drawing the tentative conclusion that food and sex as used in this experiment are approximately equal as incentives to maze learning and running by one-year-old male rats."

Food and the presence of an animal of the same kind as the subject of the experiment have been studied by two experimenters. Liddell (86) used sheep in a maze problem and offered a choice between the two above-named incentives. Ligon (87) studied rats in a maze situation and found little difference between food and the presence of other rats as incentives.

Several experimenters have complicated their studies by introducing four incentives. Kuo (82), in a multiple-choice experiment performed on rats, used an apparatus containing four compartments, one of which led to a food box by a long path, the second by a short path, the third confined the animal for a certain time, and the fourth produced an electric shock. He found that the electric shock compartment was most effective as an incentive, while the long path compartment was the least effective. Simmons (112), using groups of rats, trained them to run mazes under different conditions of incentive. The incentives used were bread and milk, sunflower seed, escape from the maze, return to the home cage, sex, litter, and bread and milk under special conditions. The results of her study showed that bread and milk, either singly or in combination with return home, sunflower seed, sex, and litter were more effective incentives than escape or return home.

From a slightly different angle, Elliott has studied the influence of appropriateness and change of reward on the maze learning of rats. In his first experiment of this sort (39), he fed his rats bran mash as reward for correct maze performance. Then, after nine days of this reward, he changed it to sunflower seed. This substitution of a new incentive for an accustomed one brought about an increase in both time and error scores. To test the influence of appropriateness of reward, Elliott (38) divided his rats into three groups. In the first group were very hungry and very thirsty rats; in the second, very hungry and slightly thirsty ones; and in the third, slightly hungry and very thirsty ones. When water was offered as a reward to the very hungry and only slightly thirsty rats, the time taken to run the maze was decidedly increased, while, in the case of the slightly hungry and very thirsty group, there was a distinct decrease in both time and error scores when the reward was changed from food to water.

A third experiment by Elliott (40), following closely along the lines of the first two, was carried out with the purpose of measuring the effect of change of incentive on maze learning. During the first nine days of the experiment, water was used as an incentive, and during the last part of the experiment, the rats were kept hungry, and food was used as a reward. The change from water to food as incentives caused an increase in time and error scores. The conclusions drawn from this experiment were that "rewards may be changed without materially affecting the learning curve, provided that the drive is changed so as to maintain an 'appropriate' relationship between drive and reward."

The influence of the strength of incentive on learning has been studied by Yerkes and Cole. Yerkes (151), in a black-white discrimination problem with the dancing mouse as subject, found that when the discrimination was easy, learning became more rapid with increase in strength of electric shock. When, on the other hand, discrimination was difficult, learning decreased as the strength of the electric shock increased. Cole (27) used chicks in a discrimination problem similar to that of Yerkes. He found that the number of trials needed to learn to discriminate correctly decreased with an increase of stimulus.

In an attempt to measure the value of red as an incentive to anger in cattle, Stratton (118) used 40 head of cattle and the experiences of 66 cattlemen. From these data he concluded that red itself is

not an incentive, but that strangeness and motion are responsible for the angry behavior of the cattle when red is used as a stimulus.

INCENTIVES FOR HUMAN SUBJECTS

The earliest studies of incentives when human beings were used as subjects were carried out in psychological laboratories. Some incentive, usually praise, electric shock, or knowledge of results, was introduced into the usual laboratory experiments to see what effect it had on the subject's performance. Interesting results were obtained, and this led the psychologist to carry his study further to see if incentives, when introduced into practical school situations, or into industry, would produce effects similar to those obtained in the more artificial situation of the laboratory experiment.

These studies, which have rapidly increased in number during the last few years, have been subdivided according to the incentives used, rather than the situation involved. They will be discussed in the following order: (a) knowledge of results, (b) praise and reproof, (c) rewards, (d) punishment, (e) the influence of an audience, (f) rivalry, (g) distraction, and (h) music.

Knowledge of Results. It is a commonly accepted fact that working blindly is very discouraging, while, on the other hand, knowledge of one's achievements acts as a spur to increased activity. How great is the incentive value of knowing what one has done, has been experimentally studied by many psychologists. The earliest studies along this line were made in connection with laboratory experiments.

One of the earliest researches along this line was made by Judd (77) in 1905. His problem was to discover "what was the effect of practice in a case when the subject was ignorant of his habit and its effect." The task consisted of judging the size and length of sides of angles. One subject was used for 10 days, and during this time he was kept in complete ignorance of his results. Judd has summarized his findings thus: "The striking fact that appears in the results of these 10 days is that practice brings little if any change. . . . There is no motive for improvement. The subject cannot see his results and cannot judge of their success or unsucess." Spencer (115) followed up Judd's work. Using 4 subjects, he found that in 3 out of 4 cases there was some improvement. He concluded that Judd's method of averaging had, to some extent, disguised the improvement which existed.

Thorndike (123) informed his subjects by using the words "right"

and "wrong" when they judged the length of strips of paper. Twenty-four subjects were used, and the results showed a decided improvement. The average percentage of reduction in error made was 61, with a *P.E.* of ± 4 . Hamilton (58, 59) used the Galton Bar in his experiment to discover the influence of incentives on judgment. The subjects were informed in the "punishment series" by the ringing of a bell when an error was made, and in the "reward series" likewise. He found that "visual discrimination of horizontal lengths by the method of average error was brought to a point of greater accuracy than had heretofore been attained."

The effect of knowledge of results on reaction-time was studied by Johanson (75). Three subjects were used and each was informed of the time taken for reacting to a sound stimulus before the next succeeding reaction occurred. The average improvement for the three subjects was 6.0%.

Sims (113) measured the influence of knowledge of results in substitution tests and rate of reading tests. The students were divided into three sections equated on the basis of the initial test scores. In the control section no incentive was given. In the group-motivated section the groups were given the average group score for the preceding practice period, while in the individually motivated section each subject was informed of his own score.

The control section showed 102.2% improvement in the substitution test, as compared with 109.9% on the part of the group-motivated section, and 157.7% in the individually motivated section. In the reading test, the control group made 8.7% improvement; the group-motivated section, 14.5%, and the individually motivated section, 34.7%. The author concluded, on the basis of his findings, that "individual motivation is vastly superior to group motivation, and group motivation is but slightly superior to no motivation other than that which comes incidentally to learning."

Féré (quoted by Burnham, 18, and Arps, 4) measured the effect of knowledge of results in ergograph experiments. Féré found that work with the ergograph decreased when the subject was blind-folded, or when the room was dark, but when it was light enough for the subject to watch his progress his work improved. Arps arranged the conditions of his experiment in such a way that each of his three subjects used the ergograph for a period of time when the results were known, and for a period of time when the results were unknown. In 38% of the work periods a higher efficiency was

shown when the results were known as contrasted with 19% when the results were unknown.

Wright (149), using the ergograph experiment, measured the effect of knowledge of results on work and fatigue. The gain in work which resulted was interpreted by Wright as due to the "mental attitudes of the subjects." Crawley (30) duplicated to a large extent Wright's study. In his incentive series, the subjects watched their progress as recorded on a smoked drum. He found that more work was accomplished when the results were visible than when they were screened.

Whiting and English (144) measured the effect of knowledge of results on a group of laboratory tests such as cancellation, multiplication, memory for nonsense syllables, steadiness, division of lines, etc. In one part of the experiment, each subject was informed of the scores made on the previous tests. It was found that knowledge of results tended to decrease the errors made, but that it did not increase the amount accomplished in all tests.

Ross (108) used the simple laboratory experiment of crossing lines to discover the incentive value of knowledge of results on amount done. One section of his subjects was given full knowledge of the work done; a second section, only partial knowledge; and the third section, no information whatsoever. The section with full information gained from 2.2% to 8.5% more than the section with only partial information, and from 4.5% to 12.6% more than the section with no information. Book and Norvell (14) studied the effect of "interest in improvement as such" in four simple learning tests with a group of college students. They found that knowledge of improvement caused a greater improvement for the group as a whole, as well as for the individual students. The group that had been making rapid and continuous improvements suddenly ceased to do so when the incentive was removed.

In practical school situations the incentive value of knowledge of results has been carefully studied. Kirby (79), using as subjects 135 school children, found that the use of graphs and other devices to acquaint the children with their success or failure in addition and division problems brought about a median gain of 48% in addition and of 75% in division during the period of his experiment. Chapman and Feder (24) followed up Kirby's study, but introduced a control group as a check on the results found. Commenting on the results of their test, the authors say: "The general trend of the re-

sults shows that the incentive exerted a considerable effect on the amount of the product."

deWeerd (141) measured the influence of knowledge of results, given in the form of class graphs or posting of individual scores, on the achievement of fifth-grade children in addition, reading, multiplication, and other similar tests. The improvement ranged from 32.3% in the case of reading to 48% in the case of multiplication. The correlation between improvement and IQ scores was .77.

To discover the influence of knowledge of success on marks in a general philosophy class, Deputy (33) divided his students into three sections, one of which had a 10-minute written review at each class meeting; the second, a 20-minute written review once a week; and the third, a 10-minute oral review at each class meeting. Knowledge of their daily success did not, according to the author, stimulate the section given the oral review to greater achievement on the final examination as compared with the other sections.

Closely associated with the incentive which comes from knowledge of success or failure is the teaching of correct and efficient methods of work as an incentive to increased output. Clark and Vincent (25) had fifth- and sixth-grade school children check columns of addition to see what effect this would have upon speed and accuracy. They found a greater efficiency in the checking group in accuracy, with a slight decrease in speed. West (142) found that remedial instruction in handwriting, with charts of sample handwriting as incentives for improvement, improved both the rate and quality of the handwriting of a group of fifth- and sixth-grade children. Symonds (119) used the Charter's Diagnostic Language Tests to stimulate a group of sixth-grade school children to improve their scores. Class charts were also used so each child could see his position in the group. The results showed that these incentives "caused learning over and above that which could be explained by practice." A similar study was made by Stone and Colvin (116) with a group of college students. Silent reading and comprehension were the two subjects of study most emphasized. The results, as measured by Stone's extension of the Monroe Silent Reading Test, showed an average gain of 74% in reading and 84% in comprehension.

Freeland (44) studied the influence of teaching on basket-ball tossing in the case of one subject. After a series of practice periods lasting for one month, an average score of 14 was made, as compared with an average score of 6.6 made during the last 10 periods

without instruction. In the field of industry, the Gilbreths (49) have applied this same principle to bricklaying, folding clothing, putting paper on boxes of shoe polish, and similar operations. They have found that the teaching of the correct method of work acts as an incentive to increased output, accompanied by decreased energy and fatigue.

Praise and Reproof. To stimulate action in human beings, praise and reproof are commonly used incentives. Back of these is the assumption that people are enormously affected by the opinions of others and that this has a great influence on their behavior. To get action, therefore, social praise or reproof have been looked upon as valuable incentives. How valuable these are, when actually tested under standardized conditions, may be seen from the following experiments.

Encouragement and praise have been the subjects of several valuable investigations. Kirby (79) encouraged his groups of school children to improve their scores and surpass their previous records, as displayed on individual charts. Median gains of 48% in addition and 79% in division were recorded. In ergograph and hand dynamometer tests with a group of 43 school children, Binet and Vaschide (10) encouraged them thus: "Allons, tu peu faire mieux que ça, toi." They found that appealing to children to break their records always resulted in improvement.

Scott (110) reported an experiment on the part of a college coach to see if encouragement caused athletes to increase their exertion. The results showed that "ordinarily the increase was marked—sometimes as much as fifty per cent." From a theoretical point of view, Ream (103) tells how to manage people in business by praise and encouragement.

Sarcasm and "razzing" have been the subjects of two experimental studies. Briggs (17) gives the results of a study of 152 high-school seniors to discover what effect sarcasm on the part of their teachers had on their work. He found that 64.5% did worse work when sarcasm was used before others, 40.7% when it was used in private, and only 17.9% when it was directed toward other students. In four tests of motor control, Laird (83) subjected a group of eight fraternity pledges to "razzing" on the part of the fraternity members. He found in all cases that steadiness was diminished, motor coordination was decreased, and rate of fatigue increased.

The relative effectiveness of praise and encouragement, as compared

with reproof and discouragement, has received much attention. The earliest study along this line was made by Gilchrist (50) in 1916 to determine the extent to which the work of college students is affected by praise or reproof. Fifty students were divided into two groups, one of which was praised and the other reproofed. The reproofed group made no improvement, in spite of practice, while the praised group improved by 79%. In 1923, Gates and Rissland (47) carried out an experiment similar to that of Gilchrist. Barnard students were given the 3-hole test and the color-naming test. One-third of the group was praised for its work, one-third reproofed, while no comment was made to the remaining third. The results showed that "encouragement was more effective than discouragement, and discouragement than mere repetition."

Hurlock (70) used the principles involved in the two above-mentioned studies and applied them to groups of school children in practical school situations. Intelligence test scores were raised as the result of praise and reproof for work done, while practice alone brought about no significant improvement. Age, sex, and race differences were discovered. The constancy of the IQ under such conditions was found by the same writer not to exist (68). Praise and reproof raised the average IQ score by 7 points, while less than one point increase resulted from practice alone.

In a later experiment, Hurlock (69) attempted to evaluate praise, reproof, and ignoring the subject as incentives to improved work. These were applied daily for one week to fourth- and sixth-grade children. The greatest amount of average improvement followed the use of praise, less following the use of reproof, and even less in the ignored group. Cohen (26) duplicated these studies, with one exception. The teacher acted as experimenter to avoid any possible incentive value which might have come in the studies previously reported when the experimenter was not the regular room-teacher. Her conclusions were that "although the use of incentives does not make a tremendous difference in the work of the class as a whole, yet incentives do seem conducive to somewhat more accurate work."

An introspective report of different incentives used by teachers was reported by Laird (84) in 1923. College students reported that public reproof decreased the amount of work they did, while public praise increased it. Similar results were obtained from reports given by high-school students (85). Briggs (16) followed up these studies with reports from 300 graduate students at Teachers' College, Co-

lumbia University, and with a study carried out in the Speyer Experimental Junior High School.

Watts (140) has discussed from a theoretical point of view the relative values of praise and reproof when applied to industrial situations.

Rewards. Little attention has been given to studying the incentive value of rewards except in industrial situations. There are several instances, however, of the study of effect of rewards on college students. Crawford (28, 29), in an extensive research program, analyzed the system of scholarship awards in use at Yale University. This study revealed the fact that the applicants for scholarships excelled the class average grades. His comment on this study is that "judging by their classroom records, the investment of University funds in these students appears to be justified."

Knight and Remmers (81) studied the reward value of possible membership in a college fraternity on work in addition. Ten freshman pledges took part in this experiment. Fifty-four juniors were used as a control group. The results showed that for the motivated group an average of 21.9 columns was added per person in each of the 5-minute periods, while for the juniors, only an average of 11. The authors explained this difference as being due to difference of motivation and of mental attitudes, rather than to difference in ability.

The value of rewards in industry, especially in the form of increased wages, has received considerable attention. Kitson (80) studied the records of output of 40 experienced hand compositors over a period of 20 weeks. A bonus was paid for all work over the standard set by the employers. An average increase of output during this period was 78%, and in all but one case, the individual output was increased. Bagby (7) reports the experience of the Gordon Vaughan Company of Dallas, Texas, where a 5-day-a-week with full 6-day salary schedule was put into operation. At the end of a year, the production increased by 35-40%, and the quality of workmanship showed a decided improvement.

A report on 600 shops of the National Metal Trades Association has been made by Odom (98). Fifty-four per cent of these used some incentive plan, and the success of it was shown not only by the fact that only 10% of the firms who used the plan gave it up, but that there was between 30 and 40% difference in the earnings of companies which used some incentive plan as compared with the straight time wage payment system.

Theoretical discussions of the value of incentives have been made by a number of writers. These are for the most part based on experimental studies of actual industrial situations or on studies of incentives made in laboratory or school situations. Bloomfield (12), Dickinson (34), Feldman (43), Hobson (62), Howard (65), Poffenberger (100, pp. 350-410), and Scott (110, pp. 10 ff.) have all made valuable contributions to the solution of this problem. Gardiner (45), from a slightly different angle, has analyzed the incentive value of the home—the reward of achievement in industrial situations.

Punishment. From the opposite angle is the use of punishment as an incentive. While it is true that in the past, punishment was more frequently used than reward, there is a gradual tendency now to lay emphasis on the value of reward as opposed to punishment. In addition to this, it is difficult to measure in experimental situations the actual influence of punishment and consequently there are few studies of this sort.

Johanson (75), in connection with his reaction-time experiment, carried out a series of tests in which the subject was told that he would receive an electric shock when he was not reacting quickly enough. The results showed that for the three subjects used in the experiment an average improvement of 14.8% occurred in the punishment series as opposed to 6.0% for the series when knowledge of results served as an incentive. Rexroad (105) followed the general principles of Johanson's work, but limited his investigation to the effect of punishment in the form of an electric shock upon inaccuracy in five problems involving choice: card sorting, pencil maze, form-substitution, cancellation, and multiple-choice. Punishment had, according to him, three effects, "disruptive, incentive and instructive." Of these three, the incentive effect displayed itself in the rapid adoption of a plan for acquiring some aid to rapid learning, and in the avoidance of errors.

Davis (32) studied the effect of corporal punishment and suspension from school. Data contained in the reports of the Board of Education of St. Louis from 1881 to 1925 were used. In spite of the increase in pupils, the use of punishment has greatly decreased. On the basis of 1000 pupils, he found that the average number of corporal punishments for the period 1881-1888 was 141.1 as compared with 1.7 for 1924-1925, and of suspensions, 10.0 in the former period as contrasted with 2.0 in the latter.

How children react to punishment was made the subject of an extensive study by Barnes (8). Data were collected from the compositions of 4000 school children in which they described what they considered to be just and unjust punishment. He found that the form of punishment that lingers longest in children's minds is some sort of bodily pain. Forty-one per cent of the group said they received punishment, even though ignorant of the reason for it. Hazard (61), from a study of 42 women prisoners at the New York State Reformatory for Women, attempted to discover whether punishment as given in the home, as well as confinement in prison, has a desirable effect on behavior. From questionnaire answers, she discovered that only 10 of the 42 women were deterred from repeating acts in violation of the law through fear of punishment, while almost 80% of the group had been subject to physical suffering for wrongdoing.

The Influence of an Audience. How the presence of others, whether as spectators or merely as co-workers, can influence the activities of people, has been the subject of many experimental studies. As is true of the majority of studies relating to the value of incentives, the first experiments to test the influence of the audience on an individual's work were laboratory studies.

Meumann (quoted by Burnham, 19) tested seven children with the ergograph and dynamometer. He found that the work was always less when they were tested alone, and that, when in the presence of the teacher alone, they did not do as much work as when they were all together without the teacher. Travis (127) tested the influence of a small audience, consisting of from four to eight upper classmen, on the work of 20 freshmen with the Koerth eye-hand coordination test. The audience was a passive one and was seated in a semicircle in front of the subject. He found that 72.7% of the subjects had their highest scores when working in the presence of an audience; 13.6% did equally well when working alone or with an audience; while 13.6% did worse when working before an audience.

Gates (46) carried out a similar experiment, using as tests the Three-Hole Test, the Woodworth-Wells Color Naming Test, a vocabulary test, and a form of the Woodworth-Wells Analogies Test. Her subjects, all Barnard students, were divided into three groups: a control group, who took the tests in the presence of the experimenter only, the "small audience group" who worked before

from 4 to 6 observers, and the "large audience group" who worked before from 27 to 37 spectators. The results showed that "no reliable differences are obtained between groups tested in the presence of but one observer and groups tested before an audience of from 4 to 6, or from 30 to 40 spectators, except in the case of the most difficult test, that of naming adjectives where there is a slight possibility of the existence of a stimulating effect due to a large audience." Burnham (18) has summarized the studies by Mayer, Meumann, Schmidt, Féré, and others relating to the influence of an audience on mental activity, and has added comments and suggestions.

Allport (1), in testing the influence of the group on association and thought, used groups of college students as his subjects. He found that the presence of a co-working group was favorable to speed of association, it produced more work in the thought test, but that a superior quality of work was obtained in the latter when the students worked alone. Following up this study, Watson (139) had 108 graduate students work at word-building tests. Part of the experiment was carried out with the students working alone, and part in groups varying from 3 to 10 persons. He found that the "product of group thinking is distinctly superior to that of the average—the larger the group, within the range of from three to ten, here studied, the more superior the group product becomes."

The influence of the group on psychological test scores was found by Weston and English (143) to be great. Out of 10 subjects, 8 did much better work when in the company of others, while 2 did work equal to that when alone. Farnsworth (42) tested the influence of the group on intelligence test scores in the case of 36 college students. The subjects took each test twice, once alone and once in the classroom. The data, according to the writer, showed no consistent or significant group effect.

Closely related to the problem of the influence of the group on mental and motor activity, is the series of studies, from a more practical point of view, of the effect of class size on the efficiency of pupils. Almack (3) has given a good summary of the studies of this sort up to 1923. Breed and McCarthy (15) measured the efficiency of pupils in classes of different sizes in terms of promotion, and found that larger classes, with an average enrollment of 45 students, surpassed smaller groups with an average enrollment of 26.72.

Thie (122), in an attempt to test the relative values of the group as opposed to the regular teaching method, divided 48 pupils into

two groups on the basis of the Thorndike-McCall Reading Test. At the end of the semester, she found that "pupils using the group method improved more in reading ability and composition work than did the class using the regular method."

Two studies have been made to test the effect of the group on solving algebra problems. Barton (9) had two equal groups of high-school students working with class discussion and with the individual assignment method. He found that, on the whole, more students from the group discussion section solved the problems than did the groups where the individual assignment method was used. Jensen and Jensen (74) have given a very extensive summary of statements of foremost educators about the suitability of class sizes for pupils of different ages, and have summarized the studies made in this field.

Griffith (54), from a slightly different angle, has measured the influence of position in a classroom on students' grades. From data collected from classes in five large audience rooms at the University of Illinois, where students are seated alphabetically, he found the average grades showed that students in the front seats have lower grades than those in the second and third rows, while those on the back row are 18.7% lower than the averages for the fourth row. The author's comment on these findings is that "variation seems essentially to rest upon the varying degrees of social integration among the members of the group."

Rivalry. The influence of the group, as a competitor, rather than as an audience, has been the subject of a number of experimental studies in recent years. Triplett (128), as early as 1897, studied the influence of pace-making and competition on bicycle races and in laboratory tests, consisting of the turning of the wheels of fishing rods. His conclusions, based on his laboratory study, was that "the bodily presence of another contestant participating simultaneously in the race serves to liberate latent energy not ordinarily available."

To see if fatigue could not be overcome temporarily by a "test incentive" Whiting and English (144) gave four subjects a series of laboratory tests, part of which were carried out without any form of incentive, and part when the subjects were divided into two competing pairs. They found no objective proof of increased work, though the subjects were convinced that they had done better because they worked harder. Whittemore (145) tested the influence of rivalry on a laboratory test consisting of hand printing in the case of 12 college students. When divided into groups competing against

each other, he found an improvement in the work of all of his subjects, though the quality of the work was poor when competition was introduced.

In substitution and rate of reading tests, Sims (113) introduced group competition as well as individual competition. The results for his tests showed that "individual motivation is vastly superior to group motivation and group motivation is but slightly superior to no motivation other than that which comes incidentally to learning." Féré (quoted by Burnham, 19), in an ergograph experiment, found that when a subject looked at a person making the same finger-movement as he the amount of work done was increased. Bykowsky (20) found that competition between groups of children "will bring forth greater effort than individual work without competition."

Moede (94), using as subjects 17 boys, aged 12 to 14 years, studied the effect of rivalry on speed of tapping and strength-of-hand-grip tests. Individual as well as group rivalry was introduced. He found that the results of the group contests exceeded those of the individual contests, while both were better than those for solitary tests. Müller (97) measured the influence of competition on arithmetic and paper-cutting tests in the case of a small group of children. He compared this incentive with that of practical utility and altruism. Competition, he found, always brought about increased work, especially among the younger children of the group.

Hurlock (71) divided 155 children of the fourth and sixth grades into "control" and "rivalry" groups, to test the influence of group rivalry on work in addition. The experiment was carried out daily for one week, and on every day the average score of the "Rivalry" group exceeded that of the "Control" group. Rivalry proved to be a greater incentive for children of inferior ability than for those of average or superior ability. Maller (88) measured the influence of competition, as opposed to cooperation, in the case of 1538 school children. He found the work done "under competition was consistently and significantly higher than under cooperation." Children of low and very high intelligence were found to be less cooperative than those of average intelligence.

Scott (110, pp. 10 ff.) has discussed in detail the influence of competition in industrial situations to show how it can be used advantageously both for the employer and the worker.

Distraction. How distractions, especially noise distractions, can serve as incentives to increased work, has been studied primarily in

laboratory experiments. Cassell and Dallenbach (21) measured the relative effects of continuous and intermittent noise distractions on reaction-time in the case of four college students. They found that distractions of this sort may inhibit or facilitate reactions, or become so habitual as to have no effect at all. Evans (41) studied the effect of light, sound, and touch distractions on reaction-time. He found that all distractions increased the length of the reaction-time. Sound proved to be the greatest distraction of the three.

Morgan (95) introduced noise distractions when a subject was working on a problem which consisted of the pressing of keys similar to typewriter keys. At first, he found that noise retarded the speed of work but later increased it. Extra effort was put forth by the subjects to overcome the distracting effects of the noises. Similar results were obtained by Dockeray (36).

Tinker (124), from a slightly different angle, measured the effect of noise sounded close to a subject's head on work involved in intelligence tests, a motor test, and two maze tests. Thirty-nine college students were used as subjects. It was found that, although the student sometimes reported a conscious disturbance from the auditory distraction, it often spurred him on to more accurate performance.

Music. The use of music as an incentive to increased work is not new. In fact, historical records show that music, especially vocal music, has been used on workers back as far as these records go. To see if music actually has any incentive value, and does increase work, several experimental studies have been carried out.

Gilliland and Moore (51) measured the effect of classical and popular phonograph selections on tapping and strength of grip in the case of 35 men students. The initial tapping record was 208 taps per 30 seconds for classical and 211 for popular music. At the end of the experiment, the difference had almost disappeared. In the beginning of the strength-of-grip tests, the popular music produced more work, while at the end of the experiment, there was a slight advantage in favor of the classical music. Diserens (35) studied the influence of music on fatigue, accuracy of movement, speed of movement, hand writing, Müller-Lyer illusion, normal suggestibility, and color selection. He found that music tends to reduce fatigue, has no definite effect on accuracy of movement, speeds up hand writing and type-writing, and reduces the extent of illusions.

The influence of music on practical situations has been the subject of three experimental studies. Ayres (6) carried out his tests during

a 6-day bicycle race held in Madison Square Garden, New York City, in 1911. One half of the distance was ridden while the band played, and one half in silence. The average time per mile with music was 3 minutes and 4 seconds, and without music, 3 minutes and 14 seconds. The average rate with music was 19.6 miles per hour, and without music, 17.9 miles. Gatewood (48) asked 56 students from the Department of Architecture of Carnegie Institute of Technology to report what effect music had on their work. Forty-nine said it made work easier. Instrumental music was preferred to vocal, and familiar music proved to be less of a distraction than music unfamiliar to the worker.

How music influences mail-sorting was measured by Smith (114) in an experiment carried out in the postoffice at Minneapolis. Classical and jazz selections were tried out on 50 night clerks. The report covered two nights with and without music. On the night when music was used, the work was completed three-quarters of an hour early, as contrasted with the no-music night, when one hour overtime for 6 men was required to finish the sorting.

REFERENCES

1. ALLPORT, F. H. The influence of the group upon association and thought. *J. Exper. Psychol.*, 1920, **3**, 159-182.
2. ———. Social psychology. Boston: Houghton Mifflin, 1924. Pp. xiv+453.
3. ALMACK, J. C. Class size and efficiency. *J. Nat. Educ. Asso.*, 1923-1924, **12**, 107-109.
4. ARPS, G. F. Work with knowledge of results vs. work without knowledge of results. *Psychol. Monog.*, 1920, **28**. Pp. 41.
5. AYLESWORTH, M. The effect of reward and punishment on visual discrimination in the white rat. M. A. thesis, Columbia Univ., 1926.
6. AYRES, I. P. The influence of music on speed in the six-day bicycle race. *Amer. Phys. Educ. Rev.*, 1912, **16**, 321-324.
7. BAGBY, N. P. Workers should be treated like customers. *Indus. Psychol.*, 1928, **3**, 104-106.
8. BARNES, E. Punishment as seen by children. *Ped. Sem.*, 1894-1896, **3**, 235-245.
9. BARTON, W. A. The effect of group activity and individual effort in developing ability to solve problems in first year algebra. *Educ. Admin. & Supervis.*, 1926, **12**, 512-518.
10. BINET, A., & VASCHIDE, E. Expériences de forces musculaire et de fond chez les jeunes garçons. *Anné psychol.*, 1897, **4**, 15-63.
11. BLODGETT, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929, **4**, No. 8, 112-134.

12. BLOOMFIELD, D. Financial incentives for employees and executives. (2 vols.) New York: Wilson, 1923.
13. BOOK, W. F. How to develop an interest in one's tasks and work. *J. Educ. Psychol.*, 1927, **18**, 1-10.
14. BOOK, W. F., & NORVELL, L. The will to learn: an experimental study in incentives to learning. *Ped. Sem.*, 1922, **29**, 305-362.
15. BREED, F. S., & MCCARTHY, G. D. Size of class and efficiency of teaching. *School & Soc.*, 1916, **4**, 965-971.
16. BRIGGS, T. H. Praise and censure as incentives. *School & Soc.*, 1927, **26**, 596-598.
17. ———. Sarcasm. *School Rev.*, 1928, **36**, 685-695.
18. BURNHAM, W. H. The group as a stimulus to mental activity. *Science*, 1910, **31**, 761-767.
19. ———. The hygiene of home study. *Ped. Sem.*, 1905, **12**, 213-230.
20. BYKOWSKY, J. Recherches experimentales sur le role de l'emulation. Warsaw, 1923. (Quoted from Maller, 92, p. 9.)
21. CASSEL, E. E., & DALLENBACH, K. M. The effect of auditory distraction upon the sensory reaction. *Amer. J. Psychol.*, 1918, **29**, 129-143.
22. CASTEEL, D. C. The discriminative ability of the painted turtle. *J. Anim. Behav.*, 1911, **1**, 1-28.
23. CHANEY, R. M., & LAUER, A. R. The influence of affective tone on learning and retention. *J. Educ. Psychol.*, 1929, **20**, 287-290.
24. CHAPMAN, J. C., & FEDER, R. B. The effect of external incentives on improvement. *J. Educ. Psychol.*, 1917, **8**, 469-474.
25. CLARK, J. R., & VINCENT, E. L. A study of the effect of checking upon accuracy in addition. *Math. Teacher*, 1926, **19**, 65-71.
26. COHEN, A. A study of incentives under schoolroom conditions. M. A. thesis, Columbia Univ., 1927.
27. COLE, L. W. The relation of strength of stimulus to rate of learning in the chick. *J. Anim. Behav.*, 1911, **1**, 111-124.
28. CRAWFORD, A. B. Effect of scholarships: a study in motivation. *J. Person. Res.*, 1925-1926, **4**, 391-404.
29. ———. Incentives to study. New Haven, Conn.: Yale Univ. Press, 1929. Pp. xii+188.
30. CRAWLEY, S. L. An experimental investigation of recovery from work. *Arch. Psychol.*, 1926, No. 85. Pp. 66.
31. DASHIELL, J. F. A quantitative demonstration of animal drive. *J. Comp. Psychol.*, 1925, **5**, 205-208.
32. DAVIS, H. H. Corporal punishment and suspension. *School & Soc.*, 1928, **28**, 623.
33. DEPUTY, E. C. Knowledge of success as a motivating influence in college work. *J. Educ. Res.*, 1929, **20**, 327-334.
34. DICKINSON, Z. C. Economic motives. Cambridge, Mass.: Harvard Univ. Press, 1922. Pp. 304.
35. DISERENS, C. M. The influence of music on behavior. Princeton, N. J.: Princeton Univ. Press, 1926. Pp. 224.
36. DOCKERAY, F. C. Attention, distraction and fatigue. *J. Comp. Psychol.*, 1922, **2**, 331-370.

37. DODSON, J. D. Relative values of reward and punishment in habit formation. *Psychobiol.*, 1917-1918, 1, 231-276.
38. ELLIOTT, M. H. The effect of appropriateness of reward and of complex incentives on maze performance. *Univ. Calif. Publ. Psychol.*, 1929, 4, No. 6, 91-98.
39. ———. The effect of change of reward on the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1928, 4, No. 2, 19-30.
40. ———. The effect of change of "drive" on maze performance. *Univ. Calif. Publ. Psychol.*, 1929, 4, No. 11, 184-188.
41. EVANS, J. E. The effect of distraction on reaction time. *Arch. Psychol.*, 1916, No. 37. Pp. 106.
42. FARNSWORTH, P. R. Concerning so-called group effects. *Ped. Sem.*, 1928, 35, 587-593.
43. FELDMAN, H. A survey of research in the field of industrial relations. In *Reports of the Advisory Committee on Industrial Relations*, Herman Feldman, Instigator. New York: Social Sci. Res. Council, 1928. Pp. 78-81.
44. FREEDLAND, G. E. Casual practice versus practice under instruction. *J. Educ. Method*, 1923-1924, 3, 203-206.
45. GARDINER, G. The home as an incentive to good work and reliability. *Indus. Psychol.*, 1928, 3, 441-446.
46. GATES, G. S. The effect of an audience upon performance. *J. Abn. & Soc. Psychol.*, 1923-1924, 18, 334-344.
47. GATES, G. S., & RISSLAND, L. Q. The effect of encouragement and of discouragement upon performance. *J. Educ. Psychol.*, 1923, 14, 21-27.
48. GATEWOOD, E. L. An experiment in the use of music in an architectural drafting room. *J. Appl. Psychol.*, 1921, 5, 350-358.
49. GILBRETH, F. B., & GILBRETH, L. M. Applied motion study. New York: Macmillan, 1919. Pp. 220.
50. GILCHRIST, E. P. The extent to which praise and reproof affect a pupil's work. *School & Soc.*, 1916, 4, 872-874.
51. GILLILAND, A. R., & MOORE, H. T. The immediate and long-time effects of classical and popular phonograph selections. *J. Appl. Psychol.*, 1924, 8, 309-323.
52. GILMORE, M. E. Coaching for intelligence tests. *J. Educ. Psychol.*, 1927, 18, 119-121.
53. GLASER, O. C. The formation of habits at high speed. *J. Comp. Neur. & Psychol.*, 1910, 20, 165-184.
54. GRIFFITH, C. R. A comment upon the psychology of the audience. *Psychol. Monog.*, 1921, 30, No. 136, 36-47.
55. GRINDLEY, G. C. Experiments on the influence of the amount of reward on learning in young chickens. *Brit. J. Psychol.*, 1929, 20, 173-180.
56. HAAS, E. L. The effect of short intervals of delay in feeding upon speed of maze learning. M. A. thesis, Columbia Univ., 1926.
57. HAMILTON, E. L. The effect of delayed incentive on the hunger drive in the white rat. *Genet. Psychol. Monog.*, 1929, 5, 137-207.
58. HAMILTON, H. C. The effect of incentives on the accuracy of judgment. M. A. thesis, Columbia Univ., 1926.

59. ———. The effect of incentive on accuracy of discrimination measured on the Galton Bar. *Arch. Psychol.*, 1929, No. 103. Pp. 74.
60. HART, H. School progress and mental test ability. *School & Soc.*, 1925, 21, 181-182.
61. HAZARD, H. H. Psychological effects of punishment. M. A. thesis, Columbia Univ., 1922.
62. HOBSON, J. A. Incentives in the new industrial order. New York: Seltzer, 1925. Pp. 160.
63. HOGE, M. A., & STOCKING, R. J. A note on the relative value of punishment and reward as motives. *J. Anim. Behav.*, 1912, 2, 43-50.
64. HOLDEN, F. A study of the effect of starvation upon behavior by means of the obstruction method. *Comp. Psychol. Monog.*, 1925-1926, 3. Pp. 45.
65. HOWARD, R. E. Reward—a powerful influence. *Indus. Psychol.*, 1927, 2, 523-525.
66. HUGHES, W. H. Why intelligence scores are not more highly predictive of school success. *Educ. Admin. & Supervis.*, 1926, 12, 44-48.
67. HUNTER, W. S. General psychology. (2nd ed.) Chicago: Univ. Press, 1923. Pp. xiii+351.
68. HURLOCK, E. B. The effect of incentives upon the constancy of the I.Q. *Ped. Sem.*, 1925, 32, 422-434.
69. ———. An evaluation of certain incentives used in school work. *J. Educ. Psychol.*, 1925, 16, 145-159.
70. ———. The value of praise and reproof as incentives for children. *Arch. Psychol.*, 1924, No. 71. Pp. 78.
71. ———. The use of group rivalry as an incentive. *J. Abn. & Soc. Psychol.*, 1927, 22, 278-290.
72. JENKINS, M. The effect of segregation on the sex behavior of the white rat as measured by the obstruction method. *Genet. Psychol. Monog.*, 1928, 3, 461-571.
73. JENKINS, T. N., WARNER, L. H., & WARDEN, C. J. Standard apparatus for the study of animal motivation. *J. Comp. Psychol.*, 1926, 6, 361-382.
74. JENSEN, M. B., & JENSEN, D. The influence of class size upon pupil accomplishment in high school algebra. *J. Educ. Res.*, 1930, 21, 120-137.
75. JOHANSON, A. M. The influence of incentive and punishment upon reaction-time. *Arch. Psychol.*, 1922, No. 54. Pp. 54.
76. JONES, H. E. Experimental studies of college teaching. *Arch. Psychol.*, 1923, No. 68. Pp. 71.
77. JUDD, C. H. Practice without knowledge of results. *Psychol. Rev., Monog. Suppl.*, 1905, 7, 185-199.
78. KEFAUVER, G. N. The life career motive and its effect on high school work. *School Rev.*, 1926, 34, 426-430.
79. KIRBY, T. J. Practice in the case of school children. *Teach. Coll. Contrib. Educ.*, 1913, No. 58. Pp. vi+98.

80. KIRSON, H. D. A study of the output of workers under a particular wage-incentive. *Univ. J. Bus.*, 1922, 1, 54-68.
81. KNIGHT, F. B., & REMMERS, H. H. Fluctuations in mental production when motivation is the main variable. *J. Appl. Psychol.*, 1923, 7, 209-224.
82. KUO, Z. Y. The nature of unsuccessful acts and their order of elimination in animal learning. *J. Comp. Psychol.*, 1922, 2, 1-27.
83. LAIRD, D. A. Changes in motor control and individual variations under the influence of "razzing." *J. Exper. Psychol.*, 1923, 6, 236-246.
84. ———. How the college student responds to different incentives to work. *Ped. Sem.*, 1923, 30, 366-370.
85. ———. How the high school student responds to different incentives to work. *Ped. Sem.*, 1923, 30, 358-365.
86. LIDDELL, H. S. The relation between maze learning and spontaneous activity in sheep. *J. Comp. Psychol.*, 1925, 5, 475-483.
87. LIGON, E. M. A comparative study of certain incentives in the learning of the white rat. *Comp. Psychol. Monog.*, 1929, 6, No. 28. Pp. 95.
88. MALLER, J. B. Co-operation and competition. *Teach. Coll. Contrib. Educ.*, 1929, No. 384. Pp. 176.
89. MANLEY, H. A power behind workers. *Indus. Psychol.*, 1928, 3, 244-246.
90. MAY, M. A. Predicting academic success. *J. Educ. Psychol.*, 1923, 14, 429-440.
91. McCABE, F. T. Effect of outside work on school success. *Lit. Digest*, Feb. 25, 1928, 96, 21.
92. McCRORY, J. A. A study of the relation between ability and achievement. *Educ. Admin. & Supervis.*, 1926, 12, 481-490.
93. McPHAIL, A. H. The correlation between I.Q. and A.Q. *School & Soc.*, 1922, 16, 586-588.
94. MOEDE, W. The effect of rivalry (Quoted from Allport, 1, pp. 280-282.)
95. MORGAN, J. J. B. The overcoming of distraction and other resistances. *Arch. Psychol.*, 1916, No. 35. Pp. 84.
96. MOSS, F. A. A study of animal drives. *J. Exper. Psychol.*, 1924, 7, 165-185.
97. MÜLLER, J. Versuch über die Einwirkung von Motiven auf körperliche und geistige Leistungen bei Schulkindern. *Zsch. f. angew. Psychol.*, 1924, 24, 81-137.
98. ODOM, W. E. Report on 600 shops of the National Metal Trades Association. *Indus. Psychol.*, 1928, 3, 228-229.
99. PINTNER, R., & MARSHALL, H. A combined mental-educational survey. *J. Educ. Psychol.*, 1921, 12, 32-43.
100. POFFENBERGER, A. T. Applied psychology. New York: Appleton, 1927. Pp. x+586.
101. POFFENBERGER, A. T., & CARPENTER, F. L. Character traits in school success. *J. Exper. Psychol.*, 1924, 7, 67-74.
102. PYLE, W. H. The relation of ability to achievement. *School & Soc.*, 1925, 22, 406-408

103. REAM, J. A tip on managing people. *J. Appl. Psychol.*, 1924, **8**, 357-361.
104. RECTOR, W. G. A study in the prediction of high school success. *J. Educ. Psychol.*, 1925, **16**, 28-37.
105. REXROAD, C. N. Administering electric shock for inaccuracy in continuous multiple-choice reactions. *J. Exper. Psychol.*, 1926, **9**, 1-19.
106. RICHARDS, A. A. Motive in education. *Ped. Sem.*, 1921, **28**, 60-72.
107. ROSENOW, C. Predicting academic achievement. *Ped. Sem.*, 1925, **32**, 628-636.
108. ROSS, C. C. An experiment in motivation. *J. Educ. Psychol.*, 1927, **18**, 337-346.
109. ———. The psychology of motivation. *Indus. Psychol.*, 1928, **3**, 347-350.
110. SCOTT, W. D. Increasing human efficiency in business. New York: Macmillan, 1923. Pp. 10.
111. SHARP, W. L. Disintegrative effects of continuous running and removal of the food incentive upon a maze habit of albino rats. *J. Comp. Psychol.*, 1929, **9**, 405-423.
112. SIMMONS, R. The relative effectiveness of certain incentives in animal learning. *Comp. Psychol. Monog.*, 1924, **2**, Pp. 79.
113. SIMS, V. M. The relative influence of two types of motivation on improvement. *J. Educ. Psychol.*, 1928, **19**, 480-484.
114. SMITH, R. D. Sorting mail to music. *Lit. Digest*, Mar. 4, 1922, **72**, 22-23.
115. SPENCER, L. T. The effect of practice without knowledge of results. *Amer. J. Psychol.*, 1923, **34**, 107-111.
116. STONE, C. W., & COLVIN, C. "How to study" as a source of motive in educational psychology. *J. Educ. Psychol.*, 1920, **2**, 348-354.
117. STONE, C. P., & STURMAN-HUBLE, M. Food vs. sex as incentives for male rats on the maze learning problem. *Amer. J. Psychol.*, 1927, **38**, 403-408.
118. STRATTON, G. M. The color red, and the anger of cattle. *Psychol. Rev.*, 1923, **30**, 321-325.
119. SYMONDS, P. M., & CHASE, D. H. Practice vs. motivation. *J. Educ. Psychol.*, 1929, **20**, 19-35.
120. SZYMANSKI, J. S. Versuche über die Wirkung der Faktoren, die als Antrieb zum Erlernen einer Handlung dienen Können. *Pflüg. Arch. f. d. ges. Physiol.*, 1918, **171**.
121. ———. Die Abhängigkeit der Lerngeschwindigkeit von der Antriebsstärke. *Pflüg. Arch. f. d. ges. Physiol.*, 1918-1919, **172**.
122. THIE, T. M. Testing the efficiency of the group method. *English J.*, 1925, **14**, 134-137.
123. THORNDIKE, E. L. The law of effect. *Amer. J. Psychol.*, 1927, **39**, 212-222.
124. TINKER, M. A. A study of the relation of distracted motor performance to performance in an intelligence test. *Amer. J. Psychol.*, 1922, **33**, 578-583.
125. TOOPS, H., & SYMONDS, P. M. What shall we expect of the A.Q.? *J. Educ. Psychol.*, 1922, **13**, 513

126. TORGERSON, T. L., & SHERMAN, I. The variability of the accomplishment of pupils of the same mental level. *J. Educ. Res.*, 1925, 11, 132-138.
127. TRAVIS, L. E. The effect of a small audience upon eye-hand co-ordination. *J. Abn. & Soc. Psychol.*, 1925, 20, 142-146.
128. TRIPLETT, N. The dynamogenic factors in pace-making and competition. *Amer. J. Psychol.*, 1897-1898, 9, 507-533.
129. TSAL, C. The relative strength of sex and hunger motives in the albino rat. *J. Comp. Psychol.*, 1925, 5, 407-415.
130. WARDEN, C. J., & AYLESWORTH, M. The relative value of reward and punishment in the formation of a visual discrimination habit in the white rat. *J. Comp. Psychol.*, 1927, 7, 117-127.
131. WARDEN, C. J., & HAAS, E. L. The effect of short intervals of feeding upon speed of maze learning. *J. Comp. Psychol.*, 1927, 7, 107-116.
132. WARNER, L. H. A study of hunger behavior in the white rat by means of the obstruction method. *J. Comp. Psychol.*, 1928, 8, 273-299.
133. ————. A study of sex behavior in the white rat by means of the obstruction method. *Comp. Psychol. Monog.*, 1927, 4. Pp. 68.
134. ————. A study of thirst behavior in the white rat by means of the obstruction method. *J. Genet. Psychol.*, 1928, 35, 178-192.
135. WASHBURN, M. F. The animal mind. New York: Macmillan, 1917. Pp. xi+333.
136. ————. Hunger and speed of running as factors in maze learning in mice. *J. Comp. Psychol.*, 1926, 6, 181-187.
137. WATSON, J. B. Behavior: an introduction to comparative psychology. New York: Holt, 1914. Pp. xii+439.
138. ————. The effect of delayed feeding upon learning. *Psychobiol.*, 1917, 1, 51-59.
139. WATSON, G. B. Do groups think more efficiently than individuals. *J. Abn. & Soc. Psychol.*, 1928-1929, 23, 328-336.
140. WATTS, F. An introduction to the psychological problems of industry. New York: Macmillan, 1921. Pp. 240.
141. WEERDT, E. H. de. A study of the improvability of fifth grade school children in certain mental functions. *J. Educ. Psychol.*, 1927, 18, 547-557.
142. WEST, P. V. Improving handwriting through diagnosis and remedial treatment. *J. Educ. Res.*, 1926, 14, 187-198.
143. WESTON, S. B., & ENGLISH, H. B. The influence of the group on psychological test scores. *Amer. J. Psychol.*, 1926, 37, 600-601.
144. WHITING, H. F., & ENGLISH, H. B. Fatigue tests and incentives. *J. Exper. Psychol.*, 1925, 8, 33-49.
145. WHITTEMORE, I. C. The influence of competition on performance: an experimental study. *J. Abn. & Soc. Psychol.*, 1924-1925, 19, 236-254.
146. WILLIAMS, K. A. The reward value of a conditioned stimulus. *Univ. Calif. Publ. Psychol.*, 1929, 4, No. 3, 31-55.
147. WILSON, H. B., & WILSON, G. M. The motivation of school work. Boston: Houghton Mifflin, 1916. Pp. 20-21.
148. WITTY, P. A., & LEHMAN, H. C. Drive: a neglected trait in the study of the gifted. *Psychol. Rev.*, 1927, 34, 364-376.

149. WRIGHT, W. R. Some effects of incentives on work and fatigue. *Psychol. Rev.*, 1906, **13**, 23-24.
150. WYATT, S. Boredom in industry. *Person. J.*, 1929, **8**, 161-171.
151. YERKES, R. M., & DODSON, J. D. The relation of strength of stimulus to rapidity of habit formation. *J. Comp. Neur. & Psychol.*, 1908, **18**, 459-482.

Columbia University
New York City

LA PSYCHOLOGIE DES STIMULANTS

(Résumé)

Dans beaucoup des études expérimentales récentes de l'apprentissage, on a concentré l'attention sur le rôle joué par les stimulants. Le terme "stimulant" s'applique aux conditions contrôlées extérieurement lesquelles stimule l'activité chez l'être humain et l'animal, opposé au terme "mobile" qui s'applique généralement aux forces dans l'individu lesquelles causent son action. Des études du rapport de la capacité et du rendement ont montré conclusivement que la plupart des individus travaillent au-dessous de leur capacité maximum, et que la plus grande différence entre la capacité et le rendement se trouve chez ceux d'une intelligence supérieure.

Les stimulants employés le plus fréquemment dans les études expérimentales de l'apprentissage chez les animaux sont ceux-ci: nourriture, eau, sexe, punition. Des comparaisons des valeurs relatives de la récompense et de la punition, de la nourriture et du sexe, etc., ont montré que la nourriture et le sexe en sont les plus puissants. On a fait aussi des études de l'intensité et de la convenance des stimulants.

Dans le cas des sujets humains, on a fait des études expérimentales dans le laboratoire psychologique, dans la salle de classe, dans l'industrie et dans les sports pour déterminer l'efficacité relative des stimulants suivants: connaissance des résultats, louange et reproche, récompenses, punition, l'assistance d'autres personnes, compétition, distraction et musique. Nul de ces stimulants n'est le plus puissant. On a trouvé que tous augmentent le rendement de l'individu, mais le degré de l'amélioration dépend de l'intensité du stimulant, des sujets employés dans les expériences et des situations où les stimulants ont été offerts.

HURLOCK

DIE PSYCHOLOGIE DER ANSPORNUNGEN

(Referat)

Bei vielen unter den neuen Untersuchungen über das Lernen hat man die Aufmerksamkeit auf die durch Anspornungen (incentives) gespielte Rolle konzentriert. Der Ausdruck 'Anspornung' bezieht sich auf *äusserlich* kontrollierte Bedingungen welche bei dem Menschen oder bei dem Tiere Tätigkeit anregen, im Gegensatz zu dem Ausdruck 'Antrieb' (motive) welcher sich vornehmlich auf *innerliche* Mächte, die die Tätigkeit des Individuums antreiben bezieht. Untersuchungen über die Beziehungen zwischen Befähigung (ability) und Leistung (achievement) haben endgültig bewiesen, dass die meisten Menschen weniger zustande bringen, als man ihrer Befähigung nach erwarten würde, und dass der grösste Abstand zwischen Befähigung und Leistung eben bei Menschen von höherer Intelligenz zu finden ist.

Die Anspornungen welche in Untersuchungen über das Lernen bei Tieren am öftersten in Gebrauch kommen sind: Speise, Wasser, sexuelle Reize, und Strafe. Vergleichen der relativen Werte von Belohnung und Strafe, Speise und sexueller Reiz, haben gezeigt, dass sich Speise und sexueller Reiz als die stärksten dieser Anspornungen auszeichnen. Untersuchungen über die Intensität und die Angemessenheit der Anspornungen sind ebenfalls gemacht worden.

An menschlichen Versuchspersonen sind experimentelle Untersuchungen angestellt worden im psychologischen Laboratorium, in Schulzimmer, in der Industrie, und im Sportfeld, zur Ermittlung der relativen Wirksamkeit der folgenden Anspornungen: Auskunft über die Resultate, Lob und Vorwurf, Belohnung, Strafe, Anwesenheit eines Publikums, Nebenbuhlerschaft, Ablenkung, und Musik. Keine dieser Anspornungen zeichnete sich als die stärkste aus. Alle, fand man, erhöhten den Ertrag des Individuums, aber es zeigte sich, dass die Besserung abhängig war von der Intensität des Reizes, von den Versuchspersonen selber, und von den Situationen in denen die Anspornungen angeboten wurden.

HURLOCK

INFLUENCE OF MOTION PICTURES ON CHILDREN'S ATTITUDES*¹

From the Psychological Laboratories of the University of Chicago

L. L. THURSTONE

This is one of a series of studies to ascertain whether the effect of motion pictures on school children can be measured and whether the effects of different kinds of pictures can be predicted. The present report concerns two experiments with the films "Street of Chance" and "Hide Out." The plan of each experiment was to ask the children to fill in certain schedules intended to reveal their attitudes on the issue which seemed to be involved in the film. These schedules were filled in by the children before and after seeing the film. If the film had any effect on social attitudes it might be revealed by changes in the scores on the schedules filled in before and after seeing the film. Such effects have been found to be measurable for several films.

The film "Street of Chance" describes the life of a gambler in such a way that the children might conceivably be affected in their attitudes toward gambling. By special arrangement this film was shown in the Strand Theater in Mendota, Illinois, on May 22, 1929. Mendota is a town of about 4000 population. There were 240 children in the experimental group in Grades 9 to 12 inclusive. A paired comparison schedule of crimes was presented to the children on May 15, 1929, before seeing the film, and also on May 23, 1929, after seeing the film. The second filling in of the schedule was done on

*Received in the Editorial Office, October 16, 1930.

¹This is one of a series of experimental studies on the effect of motion pictures on the social attitudes of children. These studies were made possible by a grant from the Payne Fund. I wish to acknowledge the assistance of Professor W. W. Charters of the Ohio State University, Mr. M. E. Steele, Superintendent of Schools, Mendota, Illinois, and Mr. O. V. Shaffer, Superintendent of Schools, Princeton, Illinois. I am especially indebted to Miss Ruth Peterson for her effective planning and managing of these experiments.

For the experiment at Mendota, Illinois, acknowledgment is also made to the Paramount Famous Lasky Corporation and to Mr. A. M. Robertson of the Strand Theater. In Princeton, Illinois, we had the assistance of the Universal Film Exchanges, Incorporated, and of Mr. Jerome Reith of the Apollo Theater.

the morning after seeing the film. The children were given free tickets to the local theatre where the film was shown in an afternoon performance. This was done in order to give as much as possible of the natural setting for the effectiveness of the picture as they are ordinarily seen by the children.

The paired comparison schedule had the following instructions:

A STUDY OF ATTITUDE TOWARD CRIME

Write your name here

Boy or girl..... Age..... Grade.....

This is a study of attitudes toward crime. You are asked to underline the one crime of each pair that you think should be punished most severely. For example, the first pair is:

speeder—pickpocket

If, in general, you think a speeder should be punished more severely than a pickpocket, underline *speeder*. If you think a pickpocket should be punished more severely than a speeder, underline *pickpocket*. If you find it difficult to decide for any pair be sure to underline one of them, even if you have to guess.

speeder—pickpocket

gambler—bootlegger

drunkard—beggar

gangster—tramp

bank robber—gambler

pickpocket—drunkard

quack doctor—bootlegger

beggar—gangster

Then followed 78 comparisons of the type indicated above. There were 13 crimes in the list and every crime was paired with every other. The children were told that they would be asked to fill in these schedules twice so that they would not be surprised when asked to do it a second time. The schedules were filled in during school hours and the picture was shown in the local theatre. Nothing was explicitly said about any connection between these two events, but there is no guarantee that some of the children may not have suspected that there was an association between them. The ideal procedure is to separate these events so that the subjects do not think of any association between the picture to which they are given free admission and the attitude schedules which they fill in during school hours.

In Table 1 we have a summary of the raw data for the schedules that were filled in before seeing the picture, and in Table 2 we have a similar table for the schedules that were filled in after seeing the picture. Table 1 shows, for every pair of offenses, the proportion of

TABLE 1

Proportion of school children in Mendota, Illinois, who said that the offense at the top of the table is more serious than the offense at the side of the table. For example, 48% of the children said that it is more serious to be a gambler than to be a drunkard. These records were made on May 15, 1929, *before* seeing the film "Street of Chance"

	b.r. 1	gam. 2	p.p. 3	dr. 4	q.d. 5	b.l. 6	beg. 7	gang. 8	tr. 9	sp. 10	pt. 11	ki. 12	sm. 13
Bank robber	1				.27	.29	.01	.50	.00	.06	.02	.73	.21
Gambler	2	.93		.05	.76	.92	.07	.92	.05	.41	.49	.90	.81
Pickpocket	3	.92		.25	.67	.75	.02	.86	.02	.39	.42	.87	.68
Drunkard	4	.95	.75		.81	.95	.01	.92	.03	.37	.62	.91	.87
Quack doctor	5	.73	.24	.33		.49	.02	.70	.02	.12	.22	.64	.55
Bootlegger	6	.71	.08	.05	.51		.00	.79	.01	.09	.26	.68	.50
Beggar	7	.99	.93	.99	.98	1.00		.96	.42	.86	.96	1.00	.99
Gangster	8	.50	.08	.14	.30	.21	.04		.02	.08	.08	.36	.31
Tramp	9	1.00	.95	.98	.97	.99	.58	.98		.91	.97	.99	1.00
Speeder	10	.94	.59	.61	.88	.91	.14	.92	.09		.58	.90	.92
Petty thief	11	.98	.51	.58	.78	.74	.04	.92	.03	.42			.78
Kidnapper	12	.27	.10	.13	.36	.32	.00	.64	.01	.10			
Smuggler	13	.79	.19	.32	.45	.50	.01	.69	.00	.08	.22	.73	

TABLE 2

Proportion of school children in Mendota, Illinois, who said that the offense at the top of the table is more serious than the offense at the side of the table. For example, 70% of the children said that it is more serious to be a gambler than to be a drunkard. These records were made on May 23, 1929, after seeing the film "Street of Chance."

	b.r.	gam.	p.p.	d.r.	q.d.	b.l.	beg.	gang.	tr.	sp.	p.t.	ki.	sm.
Bank robber		.21	.07	.05	.33	.30	.02	.50	.00	.06	.03	.62	.27
Gambler	.79		.49	.30	.64	.69	.05	.82	.04	.27	.36	.73	.69
Pickpocket	.93	.51		.30	.72	.70	.03	.87	.02	.32	.38	.84	.70
Drunkard	.95	.70	.70		.84	.87	.06	.89	.04	.33	.53	.92	.84
Quack doctor	.67	.36	.28	.16		.50	.02	.68	.01	.11	.19	.65	.54
Bootlegger	.70	.31	.30	.13	.50		.02	.73	.02	.10	.24	.70	.51
Beggar	.98	.95	.97	.94	.98	.98		.99	.36	.79	.94	.98	.98
Gangster	.50	.18	.13	.11	.32	.27	.01		.01	.06	.11	.38	.34
Tramp	1.00	.96	.98	.96	.99	.98	.64	.99		.87	.95	.99	.98
Speeder	.94	.73	.68	.67	.89	.90	.21	.94	.13		.64	.92	.89
Petty thief	.97	.64	.62	.47	.81	.76	.06	.89	.05	.36			.76
Kidnapper	.38	.27	.16	.08	.35	.30	.02	.62	.01	.08	.24		.36
Smuggler	.73	.31	.30	.16	.46	.49	.02	.66	.02	.11		.64	

the children who thought that the offense listed at the top of the table was more serious than the offense listed at the side of the table. If the proportion $P_{a>b}$ is very high, say .90, the interpretation is, of course, that the children thought generally that offense a is much more serious than b and that a should be the more heavily punished.

With these data it was possible to calculate the scale value of each of the thirteen offenses by the law of comparative judgment (2, 3, 4, 5, 6, 7). The simplest form of this law was used for these calculations, namely, Case V, so that

$$S_a - S_k = x_{ak} \sqrt{2} \quad [1]$$

in which S_a and S_k are scale values, and x_{ak} is the deviation from the mean of the probability surface in terms of its standard deviation which corresponds to the observed proportion of children who said that a was more serious than k . In the same manner we may write the equation for the two stimuli b and k in the form

$$S_b - S_k = x_{bk} \sqrt{2} \quad [2]$$

Subtracting [2] from [1], we have

$$S_a - S_b = \sqrt{2} [x_{ak} - x_{bk}] \quad [3]$$

Writing this equation in summation form so as to involve all comparisons with the stimuli a and b , we have

$$n (S_a - S_b) = \sqrt{2} [\Sigma x_{ak} - \Sigma x_{bk}] \quad [4]$$

TABLE 3
SCALE VALUES

Crime	Before	After	After (adjusted)
Gangster	.000	.000	.045
Bank robber	.006	-.069	-.028
Kidnapper	.078	.036	.082
Bootlegger	.430	.445	.512
Smuggler	.518	.423	.489
Quack doctor	.563	.429	.495
Pickpocket	1.097	.994	1.089
Petty thief	1.358	1.268	1.376
Gambler	1.536	.867	.955
Drunkard	1.610	1.344	1.456
Speeder	1.702	1.607	1.732
Beggar	2.775	2.626	2.801
Tramp	2.965	2.866	3.053

from which it follows that

$$S_a - S_b = \frac{\sqrt{2}}{n} [\Sigma x_{ak} - \Sigma x_{bk}] \quad [5]$$

This is the equation used for calculating the scale separation between the two stimuli *a* and *b*, and similar forms were used for the calculation of all other scale separations. The numerical values of Σx_{ka} and of Σx_{bk} were obtained from the tabulated proportions in Tables 1 and 2. Since the details of these calculations have been previously described, they will not be repeated here (4, 7).

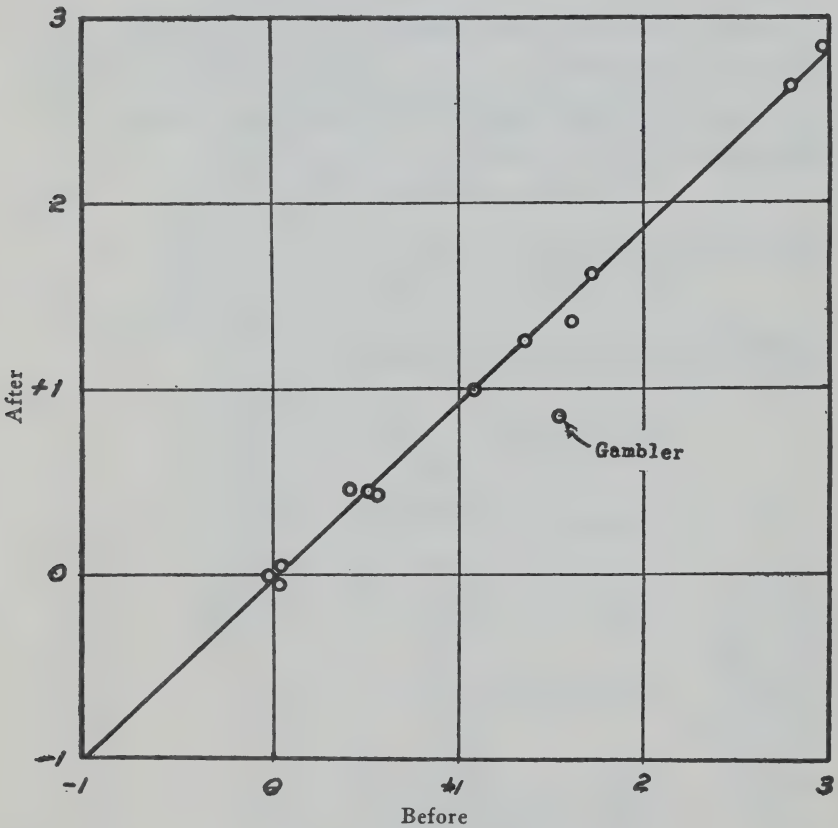


FIGURE 1

SERIOUSNESS OF CRIMES AS JUDGED BY 240 SCHOOL CHILDREN IN MENDOTA, ILLINOIS, BEFORE AND AFTER SEEING THE FILM, "STREET OF CHANCE"

The first column of Table 3 gives the names of the offenses in the list. The second and third columns give the scale values before and after seeing the film "Street of Chance." In Figure 1 we have plotted the scale values "after" against the scale values "before." A linear plot is immediately apparent with a conspicuous exception for the scale value of gambler which was evidently rated as much more serious, relative to the other offenses, after seeing the film. Low scale values represent the more serious offenses. The variations in the "before" and "after" scale values for each of the other offenses show very slight changes which may be interpreted as due to slight chance errors in the experimental proportions.

The shift in the scale value for "gambler" cannot be so interpreted. The film quite evidently had the effect of making the children regard gambling as a much more serious offense than they did before seeing the film. The same effect is shown graphically in another manner in Figure 2. Before this figure could be drawn it was necessary to adjust the scale values to a common unit. The adjustment was made on the "after" scale values. The nature of the adjustment can be explained as follows.

Imagine that the subjects were asked to fill in these paired comparison schedules one hundred times. No matter how much interest they have in the task or in the issue involved, they would get bored with the performance until their underlinings would finally become so indifferent as to be almost a chance matter. They would, nevertheless, still regard certain offenses as more serious than certain other offenses, but their indifference to filling in the schedules would obscure the affective values of the stimuli. It will be noticed that the slope of the linear plot in Figure 1 is not unity. The spread of the scale values on the second occasion is slightly smaller than the spread of scale values on the first filling in of the schedules. In fact, the slope of the line in Figure 1 is .95, as determined by the method of averages. This deviation of the slope below unity is a measure of a slight degree of indifference to the task on the second occasion as compared with the first. The discriminial error was slightly larger and the scale separations in terms of the discriminial error, therefore, seem to be slightly smaller on the second occasion. The adjustment of the second set of scale values is made by assuming that the average true scale value remained unaltered by the film and by the filling in of the schedules. A stretching factor of .95 was introduced into the second set of scale values so that the two sets would

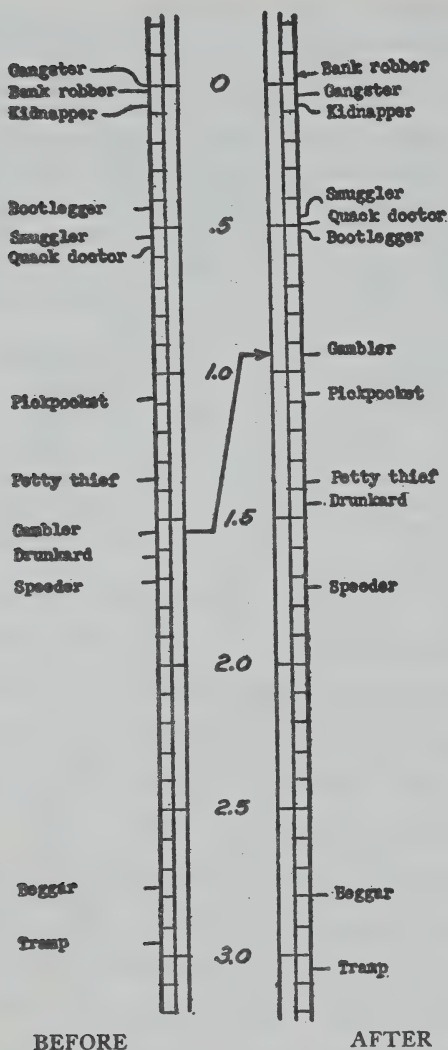


FIGURE 2

SERIOUSNESS OF CRIMES AS JUDGED BY 240 SCHOOL CHILDREN IN MENDOTA, ILLINOIS, BEFORE AND AFTER SEEING THE FILM, "STREET OF CHANCE"

be directly comparable and these two sets of scale values are shown graphically in Figure 2.

If Figure 2 were drawn without this adjustment for the slight enlargement of the discriminial error the second set of scale values would have a spread slightly smaller than the first. In the present instance, the scale value of "gambler" made such a large jump that the comparison with special regard to this one offense would be practically unaffected by the adjustment. However, theoretically, the adjustment should be made in order to take account of the fact that the discriminial error is increased by a slight amount of indifference or boredom with the repetition of the task of filling in the schedule.²

In Figure 2 we have the two sets of scale values placed in proximity for direct comparison. It can readily be seen that the scale values do not change markedly except for "gambler" which was regarded to be a much more serious offense after seeing the film than it was before the performance. The film "Street of Chance" was selected for this experiment because it was thought that it might even make the children more lenient toward gambling, owing to interest in the gambler who was the principal figure in the film. The results of the experiment show clearly that such was not the case. The film had the opposite effect, namely, to make the children regard gambling as more serious than they did before. It is also quite likely that the concept of gambling was altered in the minds of these children. It is not unlikely that many of these children were unfamiliar with high-powered gambling and that the film caused a shift in the cognitive aspects of the concept of gambling as well as in the affective aspects. On the whole, the film may be said to have a socially approved effect on the children because they became more severe in their judgment of gambling even though a gambler was an interesting principal figure in the film.

Another experiment was carried out at Princeton, Illinois, with the film "Hide Out." This experiment was planned in the same manner as the previous one. The film was shown in the Apollo Theater in Princeton, Illinois, on May 26, 1929. This town also has a population of about 4000. The film was seen by 254 school children in Grades 9 to 12 inclusive. The schedules were filled in during school hours on May 19, 1929, before seeing the film and again on May 27, 1929, the morning after seeing the film. This film was a

²This stretching factor constitutes, in effect, a measurement of boredom. This psychophysical principle may have many applications in psychological investigation.

second choice for the experiment since it was found that our first-choice film for this experiment had already been shown in this town. The experiment was completed although we were not satisfied with the film for our present experimental purposes. It involved bootlegging, and we decided to measure attitude toward bootlegging in two ways. The paired comparison schedule of offenses was given to the children before and after seeing the film. Bootlegger is included in the list of 13 offenses, and it was thought that attitudes of the children toward bootlegging might be affected by the film. We also gave this group of children a statement scale about prohibition. The instructions for this statement scale about prohibition were as follows:

A STUDY OF ATTITUDE TOWARD PROHIBITION

Write your name here

Boy or girl Age Grade

This is a study of attitude toward Prohibition. On the other side of this page you will find a number of statements expressing different attitudes toward Prohibition.

✓Put a check mark if you agree with the statement.

✗Put a cross if you disagree with the statement.

Try to indicate either agreement or disagreement for each statement. If you simply cannot decide about a statement you may mark it with a question mark.

This is not an examination. There are no right or wrong answers to these statements. This is simply a study of people's attitudes toward Prohibition. Please indicate your own convictions by a check mark when you agree and by a cross when you disagree.

Before each statement in the following list we have recorded its scale value. These scale values were, of course, omitted from the printed schedules that the children filled in. The scale values were used in determining the attitude score of each child toward prohibition. The detailed methods of constructing these scales have also been previously described so that they need not be repeated here (4).

✓Put a check mark if you agree with the statement.

✗Put a cross if you disagree with the statement.

Scale Values

- | | | |
|-----|----|---|
| 5.5 | 1. | It is absolutely immaterial whether we have prohibition or not. |
| 8.6 | 2. | The Eighteenth Amendment should be repealed and local option adopted. |

- 6.4 3. Prohibition should come as the result of education, not legislation.
- .8 4. The entire state and national resources should be mobilized for prohibition enforcement.
- 8.2 5. Liquor should be sold by licensed liquor dealers in restricted amounts.
- .4 6. Prohibition should be retained at all costs.
- 9.3 7. Prohibition is undesirable because it drives the liquor traffic underground rather than eliminates it.
- 1.4 8. Possession of intoxicating liquor in any form should subject individuals to punishment.
- 9.2 9. Prohibition should be a matter to be decided by the individual, and not by the government.
- 3.5 10. The present prohibition laws are necessary for the good of the United States.
- 7.5 11. Manufacture of wines and beer in the home should be permitted.
- 10.4 12. The open saloon system should be universally permitted.
- 6.9 13. Prohibition is not desirable now because there is not a sufficiently large majority in favor of it to make enforcement effective.
- 5.6 14. Both good and bad results have come from the Eighteenth Amendment.
- 10.2 15. Prohibition has been tried and has proved a miserable failure.
- 3.7 16. While the Eighteenth Amendment is a part of the constitution it should be observed.
- 7.0 17. Prohibition is good in principle but it is doing more harm than good because it cannot be enforced.
- 10.2 18. The Eighteenth Amendment should be repealed.
- 2.5 19. Prohibition prevents many accidents and should, therefore, be enforced.
- 4.6 20. It must be admitted that the Eighteenth Amendment is a restriction of personal liberty, but it has benefitted many people.
- 3.2 21. The national government should increase its appropriation for prohibition enforcement.
- 4.4 22. Although not completely satisfactory, the present prohibition is preferable to no prohibition.
- 3.3 23. The restriction of personal liberty under prohibition is entirely justified by the benefits.
- 4.5 24. The experiment of prohibition may prove to have some value and may, therefore, be worth trying.
- 7.8 25. Prohibition is an infringement upon personal liberty.
- 2.4 26. The effect of prohibition on the national life of America is more than constructive.
- 2.3 27. The present prohibition laws are satisfactory and their enforcement should be more severe.
- 1.8 28. Since the liquor traffic is a curse to the human family it must be dealt with by law.

These statements were adapted from a doctor's thesis by Mrs. Hattie Smith (7).

In Figure 3 we have plotted the "after" scale values of the 13 offenses against the "before" values. The diagram shows a linear plot with no conspicuous deviations, and hence we conclude that the affective judgments about these offenses were not affected by this film. The slight variations are due merely to chance fluctuations in

the experimental proportions. In Figure 4 we have plotted the frequency distributions of scores on the statement scale of attitude toward prohibition before and after seeing the film. The two means are indicated on the base lines of the distributions. This diagram reveals also no significant change in the attitudes of these children toward prohibition as a result of seeing the film "Hide Out." There was a slight change in the mean scores slightly more favorable toward prohibition after seeing the film than before seeing it, but we do not regard this change to be large enough to be attributed to the

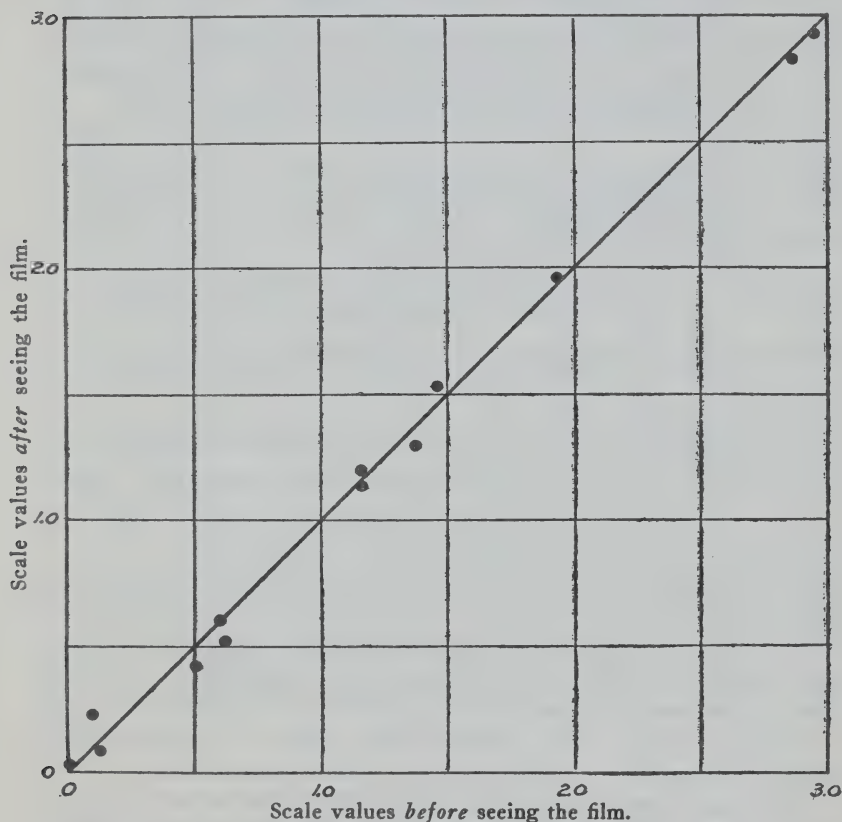


FIGURE 3

SCALE VALUES OF 13 CRIMES BY PAIRED COMPARISON FOR 254 SCHOOL CHILDREN IN PRINCETON, ILLINOIS, ON PROHIBITION BEFORE AND AFTER SEEING THE FILM, "HIDE OUT"

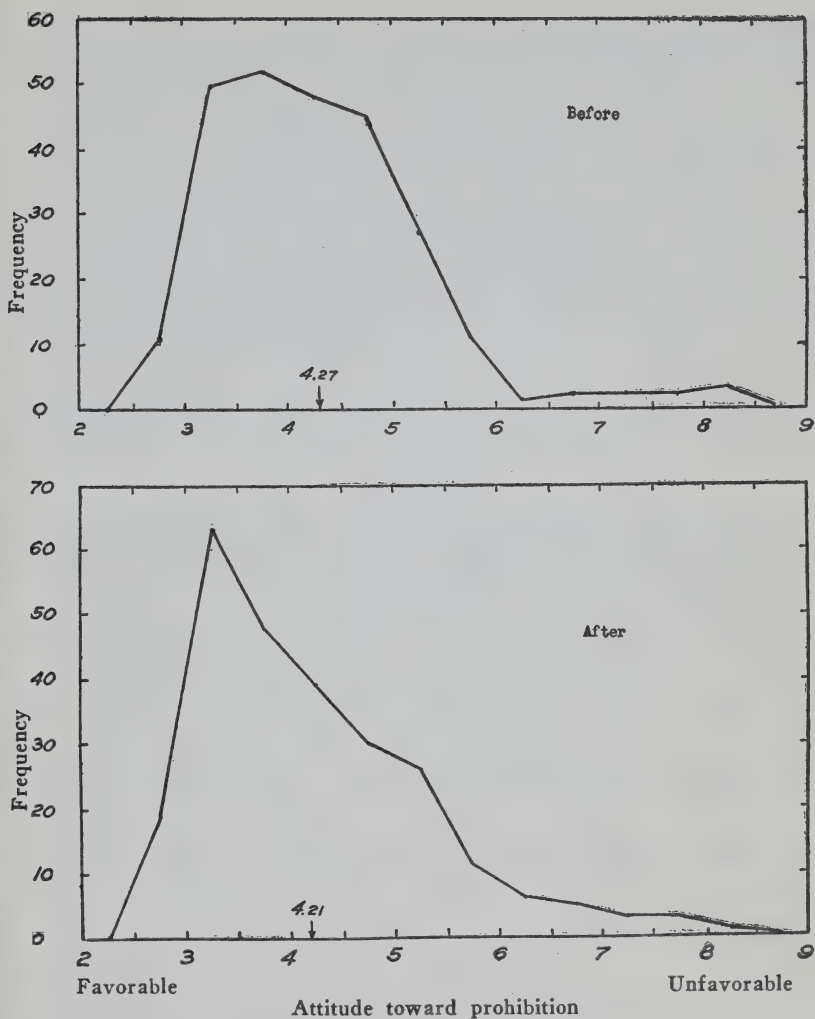


FIGURE 4

FREQUENCY DISTRIBUTION OF ATTITUDES OF 254 SCHOOL CHILDREN IN PRINCETON, ILLINOIS, ON PROHIBITION BEFORE AND AFTER SEEING THE FILM "HIDE OUT"

film with any degree of certainty. Our conclusion is, therefore, that the film "Hide Out" did not have any measurable effect on the attitudes of the children toward bootlegging or toward prohibition.

These two experiments illustrate the application of two methods of measuring affect, namely, the paired comparison procedure and the statement-scale procedure. In other experiments both of these procedures have demonstrated measurable effects of motion-picture films on the social attitudes of school children. In one of the films here discussed, namely, "Street of Chance," a conspicuous effect of the film was demonstrated. The film made the children more severe in their judgment of gambling than they were before seeing the film. It seems to be evident from these experiments and from others of a similar type that motion pictures can be used to affect the social attitudes of school children and that these effects can be objectively measured.

REFERENCES

1. SMITH, H. N. The construction and application of a scale for measuring attitudes about prohibition. Unpublished thesis, Univ. Chicago.
2. THURSTONE, L. L. The method of paired comparison for social values. *J. Abn. & Soc. Psychol.*, 1927, **21**, 384-400.
3. ———. Psychophysical analysis. *Amer. J. Psychol.*, 1927, **38**, 368-389.
4. ———. A law of comparative judgment. *Psychol. Rev.*, 1927, **34**, 273-286.
5. ———. A mental unit of measurement. *Psychol. Rev.*, 1927, **34**, 415-423.
6. ———. Attitudes can be measured. *Amer. J. Sociol.*, 1928, **33**, 529-554.
7. ———. An experimental study of nationality preferences. *J. Gen. Psychol.*, 1928, **1**, 405-425.

University of Chicago
Chicago, Illinois

L'INFLUENCE DU CINÉMA SUR LES ATTITUDES DES ENFANTS (Résumé)

Celle-ci est une d'une série d'études dans le but de constater si l'on peut mesurer l'effet du cinéma sur les écoliers et si l'on peut prédire les effets de différentes sortes de films. Ce rapport-ci décrit deux expériences avec les films "Street of Chance" et "Hide Out." Dans chaque expérience on a demandé aux enfants de remplir certaines formes faites pour révéler leurs attitudes sur le problème trouvé dans le film. Les enfants ont rempli ces

formes avant et après avoir vu le film. Si le film avait un effet sur les attitudes sociales, ce pourrait être révélé par des changements des résultats sur les formes remplies avant de voir le film et après l'avoir vu. On a trouvé qu'on peut mesurer de tels effets pour plusieurs films.

Une des formes s'est composée d'une liste de comparaisons de crimes mises en paires pour la mesure de l'attitude envers chacun de vingt crimes. Une autre forme a été une échelle pour la mesure de l'attitude envers la prohibition des liqueurs. Celle-ci s'est composée de vingt-huit opinions de la prohibition mises en échelle.

Le film "Street of Chance" a montré l'effet frappant de rendre les enfants plus sévères dans leur censure du jeu qu'avant de voir le film. Le film "Hide Out" n'a pas influé sur les attitudes des enfants envers la prohibition.

Ces expériences et d'autres de même type semblent indiquer qu'on peut mesurer les changements des attitudes sociales causés par les films et par les propagandes et les stimuli sociaux.

THURSTONE

DER EINFLUSS DES KINOS AUF EINSTELLUNGEN BEI KINDERN

(Referat)

Die gegenwärtige Untersuchung gehört einer Serie an, welche beabsichtigt, festzustellen, ob der Einfluss von Filmen auf Schulkinder gemessen werden kann, und ob die Wirkungen von verschiedenen Sorten von Filmen vorausgesagt werden können. Der gegenwärtige Bericht bezieht sich auf zwei Versuche mit den Filmen "Die Strasse des Wechselfalls" (Street of Chance) und "Der Schulpfinkel" (Hide-Out). Bei jedem Versuch verfuhr man so, dass die Kinder ersucht wurden, gewisse Protokolle auszufüllen, wodurch man beabsichtigte, die Einstellung der Kinder zu erforschen in Bezug auf die Frage mit der sich der Film zu befassen schien. Diese Protokolle wurden von den Kindern vor und nach der Besehung des Filmes ausgefüllt. Man dachte, dass wenn der Film irgend einen Einfluss auf die soziale Einstellung ausübte, dieser Einfluss sich vielleicht in Unterschieden zwischen den Ergebnissen der Protokolle vor und nach Besehung des Filmes zeigen würde. Solche Wirkungen liessen sich auch tatsächlich bei mehreren Filmen er-messen.

Eines unserer Protokolle bestand aus einer Liste gepaarter Vergleichen zwischen Verbrechen, zur Messung der Einstellung jedes der zwanzig Verbrechen gegenüber. Ein anderes Protokoll war eine Masstab zur Messung der Einstellung dem Alkoholverbot gegenüber. Dieses Protokoll bestand aus 28 rangmässig geordneten Meinungen über das Alkoholverbot.

Bei dem Film "Die Strasse des Wechselfalls" zeigte sich eine auffallende Wirkung, wodurch die Kinder in ihrer Verurteilung des Spieles viel strenger wurden, als sie vor Besehung des Filmes gewesen waren. Der Film "Schlupfwinkel" übte auf die Stellung der Kinder dem Alkoholverbot gegenüber keinen Einfluss aus.

Diese Versuche, und andere von ähnlicher Art, scheinen anzudeuten, dass durch Filme, durch soziale Einwirkungen, und durch Propaganda verursachte Änderungen in sozialen Einstellungen gemessen werden können.

THURSTONE

PROBLEM-SOLVING AMONG IDIOTS: THE USE OF IMPLEMENTS*¹

From the Training School at Vineland, New Jersey

CECELIA G. ALDRICH AND EDGAR A. DOLL

This is the third of a series of experimental studies from the Vineland Laboratory, designed to investigate (*a*) certain theories as to the nature of feeble-mindedness and (*b*) the application of the more recent methods of psychological experimentation to low-grade feeble-minded subjects.

Speculation as to the nature of feeble-mindedness leads to at least five plausible assumptions: that feeble-mindedness represents (*a*) a tendency toward lower levels of biological evolution, (*b*) a tendency toward primitive levels of human evolution, (*c*) a tendency toward permanent human infantility, (*d*) a quantitative deviation from the normal, or (*e*) a pathological variation from the normal. In any of the first four instances the feeble-minded may be said to represent very simple stages of adaptive behavior, and study of them may therefore be expected to contribute as much of psychological interest as does the similar study of normal infants, savages, or apes. Mentally deficient children and adults in institutions offer the additional advantages over these latter groups of a controlled environment and the relative absence of rapid or long-continued increments of growth. Should feeble-mindedness prove to be a pathological variation from the normal, the prevalence of distorted forms of behavior in controlled situations may throw light on normal behavior, inasmuch as it has been recognized for some time that the normal can be effectively studied by way of the abnormal.

The experimental techniques employed in previous psychological research in mental deficiency have usually required verbal comprehension and response, with consequent limitation of the choice of

*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office, October 23, 1930.

¹This is a report of a series of experimental studies of idiocy (1, 2, 3) conducted at The Training School at Vineland, New Jersey. These studies were made possible by a generous financial grant from Mrs. Leonard Elmhirst to the Department of Research. For the convenience of the reader, the theoretical premises on which the series is founded are recapitulated in the introduction to this study.

subjects to those of moron and imbecile levels of ability. On the other hand, the comparatively recent psychological experimentation with preschool children, as well as the extended experimental work with apes, has developed techniques which provide new modes of studying the low levels of feeble-mindedness in terms of comparable degrees of normal human and animal intelligence. This comparison in itself may reveal much with respect to the nature of feeble-mindedness. It may contribute a new field for the investigation of the earliest forms of intelligent behavior.

With these considerations in mind, a series of investigations has been carried forward at the Vineland Laboratory, designed to compare the adaptive behavior and genetic development of idiot children² with apes, on the one hand, and with preschool children, on the other. The ultimate program likewise contemplates a comparison between idiot subjects, under conditions of modern civilization, and primitive man. The resources of the Vineland Laboratory do not permit our undertaking such an investigation except on a library basis. It is to be hoped, however, that such a comparison might be made a part of anthropological research conducted by other investigators.

The two studies already completed have been carried forward on the assumption that idiot behavior is superior in some respects and inferior in other respects to the behavior of normal infants and of apes. It appears from *a priori* considerations that the instinctive levels of behavior are more effectively developed among animals, while adaptive behavior is probably more highly developed among idiots. Such a conclusion cannot, however, be safely reached without a fairly complete catalog of the types and degrees of behavior and the differentiation of these into early and late genetic periods. The genetic aspect of the problem has therefore been approached by comparing the reactions of idiot subjects on the verge of adolescence with those of normal preschool subjects of approximately the same degree of mental development as the idiots.

²Idiocy represents the lowest degree of mental deficiency. In terms of general ability the mental capacity of the idiot conventionally is believed to resemble that of normal children below approximately three years of age. There are, to be sure, numerous clinical varieties according to their specific etiology. Nevertheless, for present purposes, we shall consider idiocy in terms of general ability and assume that the laboratory diagnoses of our subjects are authoritative with respect to the condition of idiocy. A major purpose of this series of studies is to obtain more precise information concerning the condition.

Our first study (2) demonstrated that the technique of mental measurement developed for use with very young children is directly applicable to the measurement of mental development among idiots. The results of this study suggested the following conclusions:

1. The gross scores of idiot subjects are directly comparable with those of normal young children, as indicated by scales of genetic development. The idiot subjects are, of course, physically more mature. They obtain their scores over a wider range of test performance than do the normal preschool children, thereby revealing a wider spread of immediate capabilities.

2. Idiot subjects are definitely inferior to the normal children with respect to the development of language.

3. Idiot subjects show a corresponding nonlinguistic superiority as compared with normal children.

4. While the mental status of the idiot subject is relatively complete and static, the normal child is in process of development, with a larger potentiality for acquiring new modes and higher degrees of adaptive response.

Our second study (3) was devoted to a comparison of the adaptive behavior of idiot subjects with that of relatively mature apes. The results of this study suggested the following conclusions:

1. The application of the "problem-solving" technique employed by Köhler and by Bingham is both practical and advantageous with idiot subjects, especially since it dispenses with the language factor and puts a premium on the grosser forms of adaptive behavior.

2. The difficulty of the problem to be solved is measurable in terms of degrees of genetic development of intelligence (mental age), but individual differences other than those measured by mental age play important rôles, and incentive, or the stimulus value of the lure, is a serious factor.

3. The initial performances of idiot subjects resemble those of apes with respect to mode of reaction and the time required for successful response. It appears, however, that idiot subjects perfect their solutions more readily than do the apes, and display a comparatively greater learning facility.

In the present experiment an attempt is made to combine in a single investigation the comparative performances of apes, preschool children, and idiots, by employing a technique which has already been used comparatively for the first two groups. Unfortunately, certain important differences in the technique, necessitated by differences in the

three types of subjects, render the comparisons only plausible rather than conclusive. The study does, however, attempt to demonstrate the feasibility of bridging the evolutionary gap between the higher anthropoids, on the one hand, and the simpler levels of human development, on the other. As noted above, it would be highly desirable were comparisons also available for the primitive as compared with the infantile stages of human mental growth. Such a study makes possible not only certain vital comparisons in the relative behavior of the several groups, but also an evaluation of adaptive behavior in the anthropoids, in terms of intellectual development as measured by genetic scales of intelligence.

The technique employed for this purpose is an adaptation of that previously employed by Köhler with anthropoid subjects, and subsequently by Alpert with preschool children. Inasmuch as the technique differs in certain important particulars, the comparison of results is necessarily limited.

TECHNIQUE

The Experimental Room. These experiments were conducted in an experimental room, located in the cottage in which the subjects lived. This room measured 15 by 17 feet, with a 9-foot ceiling height. The walls and ceiling were painted a neutral shade of yellow and the floor was gray cement. The only window, which looked east, measured 4 by 5 feet, and was 4 feet from the floor of the room but level with the ground, outside. The view from this window was obstructed by a plain white curtain, fastened at top and bottom to avoid outside distractions. A diffuse electric light in the center of the ceiling supplied illumination on dark days. Ordinarily the room was sufficiently lighted by daylight alone.

The floor of the room was divided into nine approximately equal areas, by 1-inch painted lines. These areas were numbered on charts as in the accompanying diagrams (Figures 1, 2, 3, 4). This enabled the observer to make more accurate "geographic" records of the child's movements.

Throughout the experiments, the child was alone in the room. The experimenter remained outside, observing through a one-way vision screen in the door which, located at the southwest corner, permitted an unobstructed view of the entire room. Inasmuch as Köhler and Alpert were not concealed from their subjects, this constituted one important change in technique.

In the room was a pen, 8 feet square and 4½ feet high, constructed of 2-inch palings, set 4 inches apart. The pen was made in four sections, bolted together at the corners, top and bottom. Each corner was strengthened by an 18-inch brace, from the baseboard of one side to the baseboard of the adjoining side.

Köhler's anthropoids were caged, and Alpert's preschool subjects were placed in a pen which corresponded to the cage. The pen described above, however, enclosed the lure rather than the subject. It was necessary for the subject to get the lure out of, rather than bring it into, the pen. Preliminary experiments showed that idiot subjects, when placed in a pen, spent most of their time trying to get out. In such cases it was difficult to say whether this behavior represented attempts to reach the lure or to escape confinement.

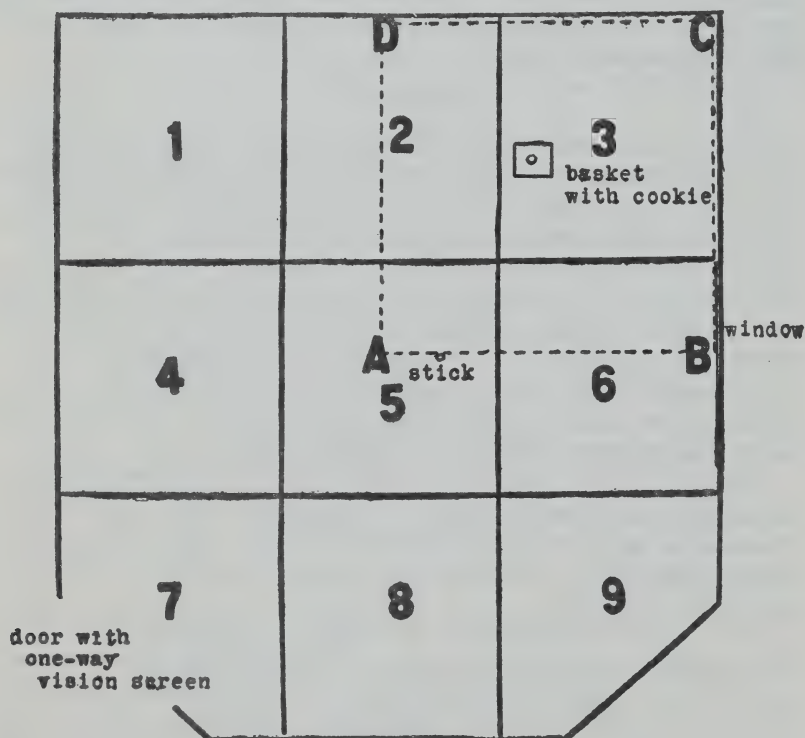


FIGURE 1
DIAGRAM OF THE SET-UP USED IN PROBLEM I

In Problem I, the pen was first placed in a corner of the room so that only two sides were exposed to the subject. Later, the pen was placed in the center of the room, so that the subject might work from any of the four sides. Inasmuch as the primary problem for the subjects of this study was to make use of sticks in overcoming the distance between themselves and a lure, all unnecessary problems in manipulation were eliminated. Allowing access to only two sides of the pen increased the necessity of skillful management of the implement, whereas, when four sides were available, the subject could either push the lure to the opposite side, or pull it toward himself. In the later problems a margin, approximately 4 feet wide, was left clear on each of the four sides of the pen.

The Problems. There were four problems, all of which involved the use of sticks as implements in obtaining a lure. Three of the four corresponded to problems in both Köhler's and Alpert's studies.

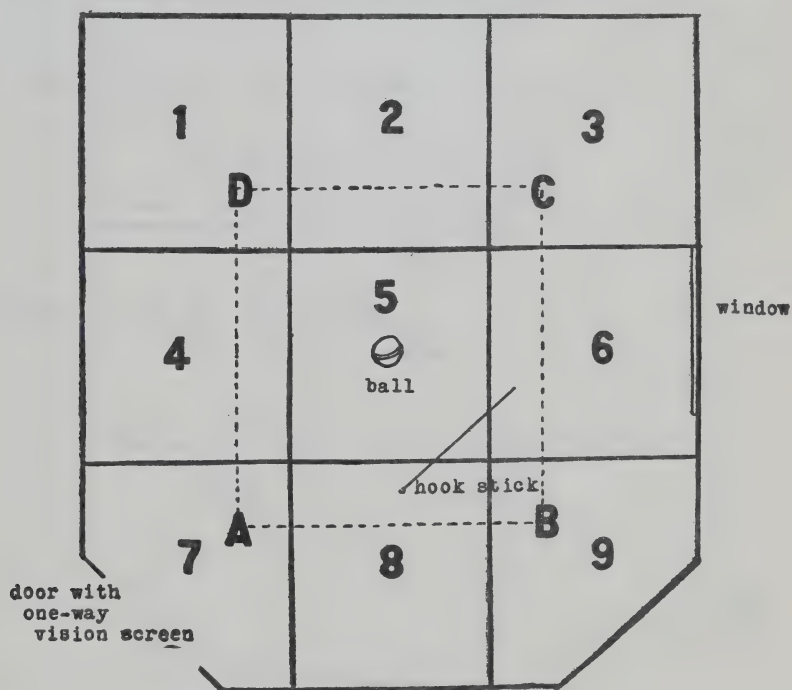


FIGURE 2
DIAGRAM OF THE SET-UP USED IN PROBLEM II

Problem I. In this problem the pen was placed in the corner of the room and a cookie, in a basket, was set in the center of the pen. A bamboo stick, 4 feet long, with a toy rake at one end, was placed upright, leaning against the third paling from the corner *A* on the side *ad*. The rake was later replaced by a stick with a wire hook at the end (7, p. 23; 4, p. 3).

Problem II. A ball, 8 inches in diameter, was used as the lure in this problem. The hooked stick lay on the floor inside the pen, across the corner *B*, and easily within reach of the subject, but not pointing toward the ball. There is no problem corresponding to this in Köhler's or Alpert's studies.

Problem III. This was identical with Problem II, except that the hooked stick was placed as far from the sides of the pen as possible, i.e., in the center position on the diagonal *AC*. Outside the

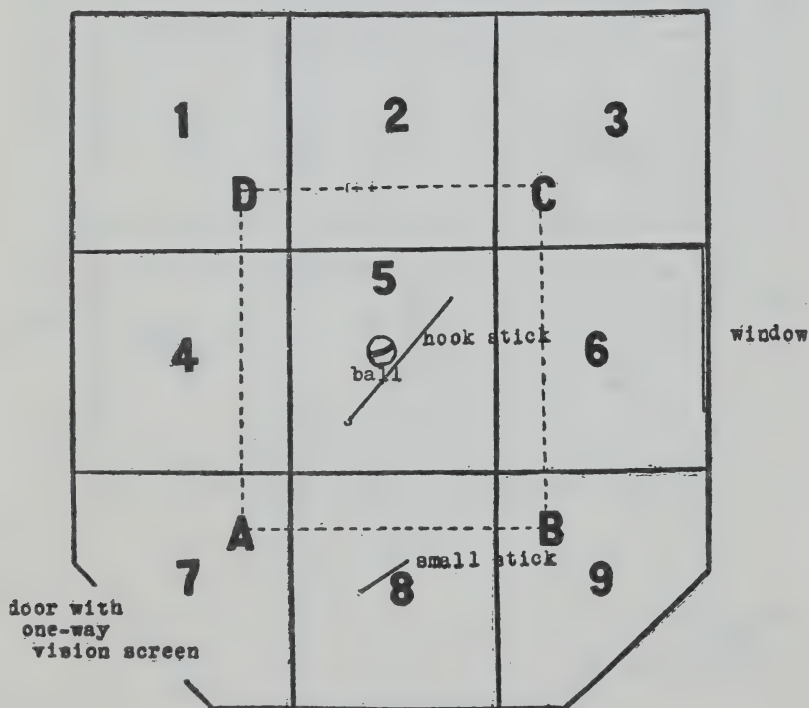


FIGURE 3
DIAGRAM OF THE SET-UP USED IN PROBLEM III

pen lay a stick, one foot in length. This stick was not long enough to reach the ball, but was long enough to reach the hooked stick, which, in turn, could be used to obtain the ball (7, p. 180; 4, p. 3).

Problem IV. In this problem the ball was placed in the center of the pen, and two bamboo sticks, each 18 inches long, were placed along the outer edge of the side *bc*. Neither of these sticks was long enough to reach the ball alone, but they were so constructed, with metal ends, that they could be joined together and used as one stick, which would be of sufficient length (7, p. 130; 4, p. 3).

The Lure. A study (1) which accompanied the previous problem-solving investigation suggested that food was probably a stronger incentive for idiot subjects than a ball or other toys. Proceeding from this suggestion, we presented a cookie in a basket as the lure

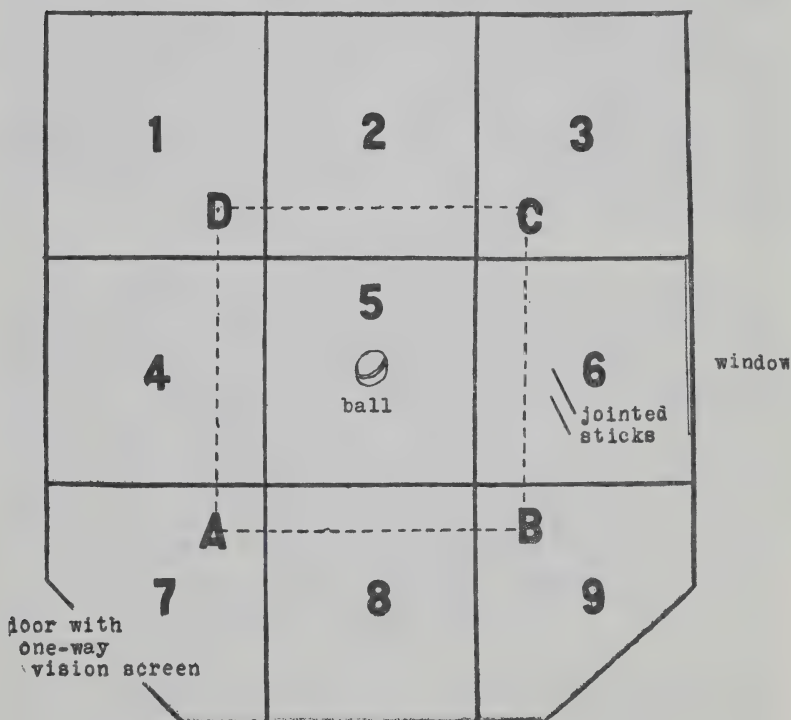


FIGURE 4
DIAGRAM OF THE SET-UP USED IN PROBLEM IV

for Problem I. The subjects of lower mentality, however, seemed to respond equally well, if not better, to a large ball, 8 inches in diameter. The ball was therefore presented in the initial trial of the other three problems. As will be seen later, the incentive value of certain lures is easily demonstrated to be subject to individual differences. However, at the present state of experimental information, it seemed more advisable to keep the incentive constant in initial presentations rather than to attempt to determine the relative value for different subjects. If, however, the subject failed several trials, a banana was substituted as a means of increasing the incentive, just as had been done in the previous box-stacking study. This change in incentive could be used with only four of the seven subjects, inasmuch as the three of lowest mentality ignored bananas, apparently failing to recognize them as food.

Another means of increasing the incentive, or, more accurately, of focusing the subject's attention on the lure, was attempted with the lower-grade subjects. If the child, after five minutes, had made no constructive effort toward obtaining the lure, the experimenter entered the room and pointed to the ball, or, as with Davey in Problems II and III, said "Get the ball." The results of this procedure will be discussed later.

THE SUBJECTS

Seven children were the subjects of this study. Although mentally and chronologically among the youngest children at The Training School, they were not idiots in the sense of the term referring to vegetative cases of feeble-mindedness. All could walk, a few could talk a little, and all were clean in personal habits and able to feed themselves. On January 1, 1930, the group ranged in life age from 8.6 to 11.9 years, and in Kuhlmann-Binet mental age from 18 to 38 months.³ The average age for standing height is given as an indication of physical development. These data are presented in Table 1.

Individual differences among idiot children are as pronounced as among normal children. Because of their specific deficiencies, the individual traits might even be said to be more pronounced. A brief

³The Kuhlmann-Binet test age was selected as a measure of general ability, primarily because of its wide use and accepted standardization. However, according to the recent study, "Comparative Intelligence of Idiots and Normal Infants," this test probably rates the subjects somewhat lower than their performance level.

TABLE 1
DESCRIPTION OF SUBJECTS

Name	LA in years	MA in months	IQ	Standing height Cm.	Age av.*
Jimmy	10.8	38	30	126	9
Carl	8.6	36	35	118	7
Buster	10.3	33	25	139	12
Sammy	8.6	32	31	128	10
Davey	11.9	20	14	128	10
Gus	10.0	20	20	129	10
Harry	10.4	19	15	129	10

*The age norms used in this study are computed from the Smedley table.

description of each subject follows, in which only outstanding differences from the group will be mentioned. Complete case history studies are on file at the Vineland Laboratory.

Jimmy was of the Mongolian type. He was eager for attention, and objected to being left in the experimental room alone. He made several attempts on each problem, but, whenever discouraged, came to the door and cried, or feigned crying. He often displayed irritability by stamping his feet and banging against the wall or door when his attempts were thwarted. He talked meagerly, in a thick voice, and with infantile articulation.

Carl, a physically under-developed boy, was quick and agile in his movements, and cared little for approval. He rarely spoke except when questioned, and then responded in monosyllables, with infantile articulation. He usually ran ahead of the experimenter into the room, but waited until the door was closed before beginning any activity toward securing the ball.

Buster was well developed physically. He had a cleft palate, and did not talk intelligibly in spite of a mental age of 33 months. He employed a relatively efficient sign language, however, by which he could make simple wishes understood. He was eager for attention and anxious to please, although somewhat deceitful as evidenced by his performance when not aware of observation.

Sammy was quick of movement, anxious to play, and well adjusted in the experimental situation. The presence or absence of the observer in the room seemed to effect no change in his behavior. His deficiency was supposedly due to a sudden arrest at two years of age, up to which time he had developed normally. On developmental

tests he showed a more even development than most low-grade children, as was brought out in the genetic study (2) of these subjects. A subjective judgment would lead an observer to conclude that Sammy's behavior more nearly resembled that of a normal two-year-old child than did that of any of the other subjects.

Davey was a Mongolian idiot, with all the physical characteristics of his type. He had the usual *hyperflexibilitas*, often accompanied by various antics and grimaces. At some time during nearly every trial he assumed a squatting position, with elbows on the floor and chin in hands, and made faces at the ball. At other times he would march about the room, stick in hand. All of his movements, especially in walking, were very awkward. When the observer was in the room Davey enjoyed teasing her, pretending to throw the ball and stopping, or avoiding catching it when it was thrown to him. He constantly teased the other children at play, slapping them or taking away their toys, apparently delighted at their discomfiture.

Gus was a pathetic little fellow, who did not talk at all and was functionally deaf. Occasionally, but not often, he responded to his name. A ball or top was his chief delight, and, although he rarely made any constructive efforts toward getting the ball, he kept his eyes on it almost constantly from the time he went into the room until he left. If the experimenter came into the room he took her by the hand and gave every indication that he wanted the ball.

Harry was a hyperactive little boy, always eager to accept any promise of change, and just as eager to return to the régime. His performances corresponded more to that which Thorndike terms random activity than did those of any of the other children. Looking at the ball never failed to be accompanied by excessive general motor activity.

PROCEDURE

The subject was accompanied to the door of the room by the experimenter, who remained outside, and closed the door after the subject had entered the room alone. As has been previously mentioned, all observations were made through the one-way vision screen in the door. Previous experience in problem-solving situations had already familiarized the subjects with the general procedure, and as a result they were all well adjusted and eager to enter the room.

Each presentation of a problem to a subject was considered one trial on that problem. If, at the end of ten minutes, the child had

not obtained the lure, that trial was considered a failure. Three successful trials was the criterion of success, and three consecutive unsuccessful trials was the criterion of failure for a single problem. The trials were always given in the morning, one each day, until the problem was either failed or solved. From three to four days elapsed between the completion of one problem and the beginning of the next. The problems were presented in the order numbered, on the assumption that this was the order of difficulty.

As in the previous problem-solving investigation, the technique required some modification. For example, the length of a trial was sometimes extended from one to five minutes if the child gave any indication of being about to solve the problem.

Another change was made in Problem I, when the pen was in the corner of the room. Carl, with the aid of the wall for support, climbed over the pen. The pen was therefore moved about 18 inches from the corner, in order to eliminate this possibility of solution.

DATA

Detailed records of each performance were made by the experimenter. These records are summarized in Tables 2, 3, 4, and 5, which list the subjects, their life ages, Kuhlmann-Binet mental ages, trial number, the position of the pen, the lure used, the result, and the time required for the solution. The subjects are arranged in the order of mental age, the subject with the highest mental age being first. The life age is given in years, the mental age in months, and the results are recorded as successful (+), failure (—), or doubtful (*D*).

It proved impossible to eliminate all solutions except one, and therefore there were occasional unanticipated solutions, which we termed doubtful, since doubt remained as to whether or not these subjects could have used the implement as planned. For example, Carl, in Problem I, insisted on climbing over the pen, thereby attaining the lure. This was effective, but difficult even at best. It was not until the possibility of such a response was removed that Carl used the stick to bring the lure to him, which is obviously a quicker and, to the adult, an easier solution. Until Carl had actually used the stick it was doubtful whether he was capable of using it as an implement.

The tables are supplemented by complete performances throughout the discussion, presented as illustrative material.

TABLE 2
SUMMARY OF PERFORMANCES ON PROBLEM I

Name	L.A. in years	M.A. in months	Trial No.	Position of pen	Implement	Lure	Result	Min.	Time Sec.
Jimmy	10	38	1	corner	rake	cookie	+	10	
			2	"	"	"	+	1	
			3	"	hook	"	+	4	
Carl	8	36	1	corner	rake	cookie	D	10	
			2	"	"	banana	D	1	
			3	18" from wall	"	cookie	+	3	
			4	corner	"	"	+	3	
			5	"	"	"	+	2	
Buster	10	34	1	corner	rake	cookie	+	2	
			2	"	"	"	+	1	20
			3	"	hook	"	+		
Sammy	8	32	1	corner	rake	cookie	+	2	45
			2	"	hook	"	+		45
			3	"	"	"	+		
Davey	11	19	1	corner	hook	ball	+	1	
			2	"	"	"	+	2	
			3	"	"	ball in basket	+	4	
Gus*	10	19	1	corner	hook	ball	—	10†	
			2	"	"	"	—	10	
Harry	10	18	3	"	"	"	—	10	
			4	"	"	"	—	10	
			1	corner	hook	ball	—	10‡	
			2	"	"	cookie	—	10	
			3	"	"	"**	—	10	

*Problem II was presented between Gus's second and third trials.

†The experimenter entered the room after 5 minutes and pointed to the ball.

‡The experimenter entered the room after 8 minutes and pointed to the ball.

**During this trial the cookie was placed closer to the subject than had been the case previously.

TABLE 3
SUMMARY OF PERFORMANCES ON PROBLEM II*

Name	L.A. in years	M.A. in months	Trial No.	Position of pen	Implement	Lure	Result	Min.	Time Sec.
Jimmy	10	38	1	center	hook	ball	+		25
			2	"	"	"	+		20
Carl	8	36	1	center	hook	ball	+		30
			2	"	"	"	+		20
Buster	10	34	1	center	hook	ball	+	1	30
			2	"	"	"	+		
Sammy	8	32	1	center	hook	ball	+		45
			2	"	"	"	+		30
Davey	11	19	1	center	hook	ball	—	13†	
			2	"	"	"	—	10	
			3	"	"	cookie	—	10	
Gust‡	10	19	1	center	hook	ball	+	2	
			2	"	"	"	+	3	
			3	"	"	"	+	9	
Harry	10	18	1	center	hook	ball	—	10	
			2	"	"	"	—	15**	
			3	"	"	"	—	10	

*This problem was solved immediately by four subjects, their solutions being so perfect that it was considered unnecessary to give them more than two trials.

†The experimenter entered the room after 10 minutes and said, "Get the ball."

‡Problem I was presented between Gus' second and third trials.

**The experimenter entered the room after 10 minutes and said, "Get the ball."

TABLE 4
SUMMARY OF PERFORMANCES ON PROBLEM III

Name	L.A. in years	M.A. in months	Trial No.	Position of pen	Implement	Lure	Result	Min.	Time Sec.
Jimmy	10	38	1	center	hook and hammer stick	ball	+	2	30
			2	"	hook and small stick	"	+	1	
			3	"	"	"	+		
Carl	8	36	1	center	hook and small stick	ball	+	1	30
			2	"	"	"	+	1	30
			3	"	"	"	+		
Buster	10	34	1	center	hook and hammer stick	ball	—	10	
			2	"	hook and small stick	"	+	1	30
			3	"	"	"	+		45
			4	"	"	"	+		
Sammy	8	32	1	center	hook and small stick	ball	+	2	30
			2	"	"	"	+		30
			3	"	"	"	+		
Davey	11	19	1	center	hook and small stick	ball	—	10*	
			2	"	"	"	—	10	
			3	"	"	"	—	10	

*The experimenter entered the room after 5 minutes and said, "Get the ball."

TABLE 5
SUMMARY OF PERFORMANCES ON PROBLEM IV

Name	L.A. in years	M.A. in months	Trial No.	Position of pen	Implement	Lure	Result	Min.	Time Sec.
Jimmy	10	38	1	center	jointed stick	ball	—	10	
			2	"	"	"	—	10	
			3	"	"	banana	—	10	
Carl	8	36	1	center	jointed stick	ball	+		15
			2	"	"	"	+		30
			3	"	"	"	+	1	15
Buster	10	34	1	center	jointed stick	ball	+		15
			2	"	"	"	+	2	
			3	"	"	"	+		15
Sammy	8	32	1	center	jointed stick	ball	D		
			2	"	"	"	—	10	
			3	"	"	banana	+	1	30
			4	"	"	ball	+	1	
			5	"	"	"	+	1	
Davey	11	19	1	center	jointed stick	ball	—	10	
			2	"	"	"	—	10	
			3	"	"	"	—	10	

DISCUSSION

The four problems used in this study were designed to display varying grades of difficulty. It was expected that, while the first might be solved with relative ease, the last would present difficulties to all the subjects. Each problem added one more complicating factor to the situation, and therefore, by consideration of the steps at which the failures occur, we should be able to observe the formative period of a very simple response. As will be evident, our pre-suppositions as to what constituted a difficulty for idiot subjects were not always correct.

Types of Response. For purposes of discussion, we have divided the types of response into three groups, which include only the initial trials on each problem. Such a limitation to initial trials eliminates the influence of special training, perseveration, and latent learning from this phase of the discussion. Table 6 presents the time and result of each first trial, in its respective group.

Group 1 includes those immediate solutions in which no problem seemed to exist. The child solved without hesitation, trial and error, or random activity. According to adult standards, these responses constituted the perfect logical solution, that is, the child entered the room, looked about, picked up the implement and used it as a means of obtaining the lure. We can say without hesitancy that these are the highest types of performance, inasmuch as, according to Table 6, they invariably resulted in success; there were no errors, and the time required for success with this type of response was in no case over two minutes. Eight of the 24 initial trials fell in this group.

Group 2 includes those performances in which there was a certain amount of exploration prior to solution, and two or more unsuccessful attempts to obtain the lure. These attempts might have been either logical or absurd from the observer's point of view, but they were all centered about the lure. Examples of this type of behavior were: reaching with one stick when two were necessary, attempting to climb over the side of the pen, trying to squeeze between the palings, or even calling to someone to get the lure. The results included both failures and successes, and the time ranged from one to ten minutes. Ten of the 24 first trials fell in this group.

Group 3 includes those trials made up primarily of irrelevant activity, that is, the child performed some act or acts which apparently

TABLE 6
INITIAL PERFORMANCES CLASSIFIED ACCORDING TO TYPE OF RESPONSE

Subject	Problem	Group 1 Time		Result	Problem	Group 2 Time		Result	Problem	Group 3 Time		Result
		Min.	Sec.			Min.	Sec.			Min.	Sec.	
Jimmy	II		25	+	I	10		+	I			
					III	2	35	+				
					IV	10		-				
Carl	II		30	+	I	10		D				
	IV		15	+	III	1		+				
Buster	II	1		+	I	2		+	III	10		-
	IV		15	+								
Sammy	II		45	+	I	2		+				
	III	2		+	IV	10		D				
Davey	I	1		+					II	13		-
Gus									III	10		-
									IV	10		-
Harry					II	2		+	I	10		-
					II	10		-	I	10		-

had no connection with the solving of the problem. This did not include activity which, although irrelevant from the observer's point of view, may have appeared relevant to the child, such as crying for someone to come, or banging on the pen. Relevancy was judged by whether or not the activity appeared to be centered about the lure. Under irrelevant responses were included pounding on the walls for amusement, marching about the pen and scraping the rake along the floor, etc.⁴ The trials of this group were all failures, and included 6 of the 24.

The above classification into types of solution is largely a matter of convenience. However, the ease with which these performances lend themselves to such a classification indicates that the division is not unjustified. Considering the time taken and the result, each group yields a certain amount of conformity, making responses of Group 1 superior and those of Group 3 inferior. They will be considered as such throughout the discussion.

Comparison of Results of the Four Problems. Accepting the responses of Group 1 as superior to those of the other two groups, we might infer that Problem II was easier than any of the other problems. It was solved immediately by four subjects, whose solutions were so perfect on both first and second trials that the third trial was considered unnecessary. This relative ease is further confirmed by Gus, who made a Group 2 solution on this problem after having completely failed Problem I. This made five solutions on Problem II and two failures. The following account will show that neither of these failures was complete.

Davey, using the hook, pushed the ball around but scored a failure, inasmuch as he did not remove it from the pen. The following complete account of Harry's performance reflects the secondary difficulty of lifting the ball over the pen rather than using the hook to bring it within reach.

Subject: Harry

Problem II—Implement: hooked stick. Lure: ball. *Trial 1.* Walked around pen, looked in, stooped, took hold of stick and let go immediately. Continued to walk around room and finally came to door. Left door and, going to wall of block 4, sat

⁴Irrelevant activity might, of course, have been included in a trial which was primarily of the Group 2 type. In these cases the performance was listed in the group which was characteristic of more than half of the behavior involved.

down (3 to 5 minutes). Got up and came to door. Left door and went to pen at *bc*, took stick and pushed the ball to the side *ab*. Started to lift the ball over, but dropped it. Stooped and pushed it away, then went to wall of block 8 and sat down (6 minutes). Jumped up, ran to door and then around the room. Activity increased, with much running about and vocalizing. The experimenter entered and said "Get the ball, Harry." Harry picked up the ball and again dropped it. Tried to go between palings. Came back to door and then returned and again tried to get between palings. Time: 10 minutes.

Problem I also had five solutions, but only one was immediate, whereas four were of the Group 2 type. The two failures were of the Group 3 type, displaying only irrelevant activity. This inferiority of performance might be attributed to the fact that Problem I was an introduction to the entire series and led to better solutions on Problem II as the result of experience on Problem I. This factor undoubtedly had some influence, the extent of which it is very difficult to estimate without a control group. However, if experience were the influential factor, Gus should have been able to solve Problem I after having solved Problem II. After two successful trials in Problem II, he was therefore allowed two more trials on Problem I, with the pen in the center of the room so as to allow four sides free from which he might work. His performances were, as before, entirely irrelevant. He played with the stick, but did not once put it through the palings toward the ball. He looked at the ball but gave no other indication of being aware of it. Problem II was again presented, and at the end of 9 minutes he used the hook and removed the ball from the pen.

This suggests that the proximity of the implement to the lure played an important rôle in the solution. In Problem II the subject could see both the lure and the implement at the same time, while in Problem I the implement stood upright against the outside of the pen. Köhler describes a similar phenomenon occurring with apes. They could use an implement placed between themselves and the lure, but when the implement was behind the ape the problem became more difficult.

The performances of Jimmy and Buster, in Problem I, tended to confirm the above suggestion. Jimmy tried various methods, such as climbing over the pen, lifting it, and unfastening the bolts. Finally, he became interested in the rake *per se*. The first time he

looked at the lure, with rake in hand, he immediately stopped what he was doing (scraping the floor with the rake) and used the rake to draw the lure toward him. Buster merely looked over the pen and then around the room. He finally picked up the rake and began scraping the floor, then returned to the pen and again looked over. Suddenly he looked at the rake in his hand and immediately used it. In these two performances it would seem that the rake was not seen as an implement until it was in the subject's hand, thereby bringing about a greater proximity to the subject and the lure.

Judging from the results, Problem III was more difficult than either Problems I or II. Gus and Harry were not given Problem III, inasmuch as success required that the subject perform the same solution called for in Problem II, with the additional task of using the small stick to bring the hook within reach. Davey failed this problem much as he failed Problem II. He obtained the hook and pushed the ball, but did not remove it from the pen.

During the first presentations of Problem III a small stick, with a short handle at right angles to the stick, was used. This gave it a mallet-like appearance and was to be used to pull the hook within reach. The "hammer stick," as it was named by Jimmy, proved to be too much of a distraction, however. Jimmy hammered the palings, the hook, and, finally, the ball, but at the same time continued with the solution. Buster, however, hammered to the exclusion of all other activity. As a result of these two trials, a small plain stick was substituted, which was far less a distraction and served the purpose equally well. The first trials on this problem resulted in one immediate solution, by Sammy, two solutions after several attempts, and two failures made up of irrelevant activity. The two responses of the trial-and-error type were almost immediate solutions. Both Carl and Jimmy made only one attempt prior to the real solution, and that was to reach toward the ball with the small stick, before getting the hook.

The results on Problem IV indicated increasing difficulty. Since Davey had approached the solution on Problems II and III, he was presented with Problem IV, which he failed completely, making no constructive effort whatever. Of the remaining four subjects, Carl and Buster made immediate solutions, Sammy solved when a banana was substituted for the ball, and Jimmy failed after many constructive attempts.

The majority of first responses, as has been observed, are of the

Group 2 type (trial-and-error). This was especially true of Problem I, which was the introduction to the entire series. Analysis of the individual performances also reveals the fact that the trial-and-error solutions of Problem I included many more attempts than did the same type of solution for Problems III and IV, although the latter were more difficult to solve. It is assumed, therefore, that experience on Problem I helped to eliminate some of the absurd possibilities which might otherwise have been tried on the later problems.

From this discussion we might conclude (*a*) that these four problems are graded in difficulty, Problem II being easier than Problem I, with Problems III and IV following in numerical order; (*b*) that greater proximity of the implement to the lure or subject is accompanied by an increase in the number of successes; (*c*) that implements which are attractive *per se* may inhibit the solution, while (*d*) change of incentive may facilitate the solution; and (*e*) that the trial-and-error type of solution is more frequent in the first trial of Problem I than in any of the other problems, and is also more frequent in the total number of first trials than is any other type of solution. This suggests that the characteristic response of idiot subjects in a new situation is a process of exploration, with a number of attempts, some logical and some foolish, until the correct solution is found.

Learning in Problem-Solving Situations. Inasmuch as three solutions were required as a criterion of success on all except Problem II, there is an opportunity to measure the first results of learning in this type of situation. For the purpose of this discussion, learning will be arbitrarily defined as the acquisition of a new response and the improvement of it as measured by time. By "new response" we refer to the ability to solve the problem which, while new with reference to the total situation, may or may not be new in its details. The time of each solution was considered as the cumulative time of trials up to and including the solution. Thus, one ten-minute failure prior to a one-minute solution is equivalent to one solution of eleven minutes.

Time, as a criterion of learning, involves obvious inconsistencies inasmuch as it ignores so much of the purely qualitative reaction. However, from the standpoint of effectiveness, it is the most adequate and objective in a situation where the errors cannot be controlled or satisfactorily weighed. Since there were only a few subjects, especially after eliminating those who failed, averages are of no con-

sequence. However, the data on individuals, presented in Table 8, indicate tendencies worth noting.

In Problem I there were five subjects, in Problem III four subjects, and in Problem IV only three subjects. With only four exceptions, the greatest drop in time occurred between the first and second solutions, with little, none, or possibly an increase, in time between the second and third. The solutions, as has been noted, were primarily of a trial-and-error type. The majority of errors were eliminated, however, between the first and second trials. This same type of behavior has been observed among apes and is presented by Köhler as evidence of "insight."

Although the similarity of performance is recognized, at this stage of experimental investigation we cannot attribute "insight," as commonly defined, to idiot children. Nor can we conclude that such behavior is due to learning by the arousal of configural patterns, even though the same behavior among apes has led to that conclusion. In fact, causal explanations of any kind are avoided, allowing the merits of this investigation to be derived from the recognized similarity in observed behavior of the idiot child and the ape.

TABLE 8
TIME REQUIRED FOR SOLUTION ON EACH TRIAL

	Solu- tion No.	Jimmy		Carl		Buster		Sammy		Davey	
		Min.	Sec.	Min.	Sec.	Min.	Sec.	Min.	Sec.	Min.	Sec.
Problem I	1	10		14		2		2		1	
	2	1		3		1		45		2	
	3	4		2			20	45		4	
Problem III	1	2	30	1		11		2			
	2	1		1	30		30		30		
	3		30		30		45		30		
Problem IV	1				15		15	11	30		
	2				30	2		1			
	3			1	15		15	1			

Success and Mental Age. The plus and minus scores in Tables 2, 3, 4, and 5 reveal the fact that the first failure, as well as the greatest number of failures, occurs among the three lowest-grade subjects, Davey, Gus, and Harry. There is, however, considerable difference between the performances of these three, which suggests, as in the box-stacking study, that factors other than mental age as

measured by the Binet scale play an important rôle in success and failure. This fact involves an important similarity between our subjects of this low level of intelligence and their superiors, among whom such factors as physique, personality, experience, and interests are recognized as playing important rôles in success or failure.

Among the three subjects of lowest mental age, Harry failed all the problems, although he partially solved Problem II; Gus solved Problem II but completely failed Problems I and III; while Davey gave a perfect solution on Problem I, partially solved Problems II and III, but completely failed Problem IV. A subjective judgment on Davey's performance in Problems II and III would be that he could easily have completed the solution, but lacked a sufficiently strong incentive. This, of course, would place his performances on a qualitatively higher basis than those of either Gus or Harry. It is therefore evident that even among these three the individual abilities varied. Harry was quite excitable, and, although activity rarely ceased while in the experimental room, it consisted almost entirely of random movement. Gus, on the other hand, was more apathetic and stood looking at the ball throughout trial periods, rather than going distractedly from one thing to another, as did Harry. Davey was very stable, slow of movement, and persistent in his efforts if the incentive was sufficiently strong. It is possible that these personality differences may partially account for the difference in performance ability not measured by developmental scales.

The mental ages of the remaining four children ranged between 32 and 38 months. This difference is within the probable error of the Binet scale, and these subjects may therefore be considered of approximately the same mental age. The number of trials, together with the number of successes, failures, and doubtful solutions of each

TABLE 9

NUMBER OF SUCCESSES AND FAILURES OF FOUR SUBJECTS WITH MENTAL AGES FROM 32-38 MONTHS

Name	M.A. in months	No. of trials	No. of successes	No. of failures	No. of doubtful solutions
Jimmy	38	10	7	3	
Carl	36	12	10		2
Buster	34	11	10	1	
Sammy	32	12	10	1	1

child in this group, are listed in Table 9. Among these results the varied effect of incentive and distractions on four subjects of approximately equal mental age and physical development is displayed.

Three of these four subjects with equal mental ages solved all four of the problems, but Jimmy, who actually tests six months above Sammy, failed Problem IV completely. Sammy succeeded on Problem IV only with an increase in incentive, and Buster succeeded on Problem III only after a distraction had been removed. Carl and Buster immediately solved Problem IV, which Jimmy could not solve at all, despite repeated attempts and an increase in incentive.

From these examples we might conclude (*a*) that the subjects of lowest mental age fail most frequently, but (*b*) that other factors than mental age are involved, inasmuch as there is variation in the degree of success attained by subjects of approximately the same mental age.

Variation in Incentive. The incentive was varied in three ways: (*a*) by substituting a banana for the ball, (*b*) by pointing to the ball or saying "Get the ball," and (*c*) by removing distracting elements.

The first variation was tried on two occasions, both in Problem IV. Jimmy had failed two trials in spite of many and varied attempts to reach the ball. He had used first one stick and then the other, he had lain prone with arm stretched through the palings in order to have the benefit of its maximum reaching length, and, finally, becoming quite discouraged, he had even come to the door and cried. For the third trial a banana was substituted for the ball, and he repeated all of his former attempts. Not once did he show any indication of joining the sticks, even though he held them together as if to measure and determine which was the longer.

Sammy, on the other hand, had managed to get the ball with one of the pair of sticks on his first trial. He repeated his efforts on the second trial and failed. The banana was substituted for the lure in the third trial and in less than two minutes Sammy joined the sticks and obtained the lure. On the fourth and fifth trials the banana was replaced by the ball, which made no noticeable difference in Sammy's response as given on the third trial.

The second means of reinforcing the incentive, by pointing to, or telling the subject to get the ball, was used only with the three lowest-grade subjects. In every case this added stimulation increased the activity. Harry tried to slip between the palings, Davey picked

up the stick, and Gus looked and jumped about, but in no instance did this variation in incentive replace failure by success.

The third method, that of eliminating distractions, was tried twice, (*a*) by substituting the hook for the rake, and (*b*) by replacing the small stick for the so-called "hammer stick." The former substitution brought about a more immediate response, with less play interrupting the solution, and the latter substitution brought about a success in place of a failure. Buster had played during the whole trial with the "hammer stick," and at the beginning of the next trial made signs indicating that he wanted it. When it was not available he immediately obtained the lure.

All of the changes, whether substitution of lures, pointing to the ball, or elimination of distractions, seemed to focus the child's attention on the lure. In the two instances where a substitution was made, the subjects made more constructive efforts; in the two instances of pointing to the ball, the activity of the subjects increased; in the two instances of eliminating distractions, activity was centered about the lure only.

From these results we may conclude (*a*) that any change in incentive which tends to focus the subjects' attention on the lure will facilitate the solution of the problem, but (*b*) that the extent to which such changes will be effective are, broadly speaking, dependent upon the child's general level of ability.

Comparison of Idiot Behavior with That of Preschool Children and Anthropoids. Inasmuch as Köhler's study dealt with 9 apes, Alpert's with 44 preschool children, and ours with only 7 idiot subjects, a quantitative comparison of results is not feasible. An analysis of qualitative reactions, however, yields many interesting points of similarity among the three groups, although even these must be accepted cautiously, due to differences in the technique of presenting the problems.

We can dismiss the quantitative results with the statement that these problems appear to be at the borderline of difficulty for all three groups. Some of the preschool children, apes, and idiot subjects solved all of the problems, while some of each group failed. Alpert found a very low correlation between the child's mental age on the Stanford-Binet and the ability to solve the problems. This suggests that among preschool children, as among idiot children, other factors than mental age play an important rôle.

The three groups displayed interesting similarities in modes of

attack. The most common first response for all groups was reaching. Very few subjects of any group responded with entirely random activity. According to Alpert, such activity was more common among apes than among preschool children, and, as nearly as can be determined, the idiot child probably falls between the two groups. Gus and Harry very definitely responded in this manner.

The trial-and-error responses included in our Group 2 correspond to Alpert's "exploration of elimination" response. This type of activity is the most common among the three groups.

Both Köhler and Alpert attribute "insight" to their subjects and consider it responsible for the final solution. Although the observed behavior of our subjects is very similar to that of their groups, we hesitate to attribute it to "insight," or, in fact, to make any inferences with respect to cause.

The idiot subjects, like the apes and preschool children, were aided by the elements of the situation being brought into greater proximity with each other, either by the subject or by the experimenter. An implement between the subject and the lure was used more readily than one behind the subject, and if, by chance, the child held the implement in his hand the connection between it and the lure seemed more easily made.

There was evidently some transfer from one problem to the next, inasmuch as Problem III, which was merely a complication of Problem I, was solved in less time on the initial presentation than was Problem I. Such transfer was noted also among apes and preschool children.

The temperamental and emotional factors found among preschool children are present to a certain degree among idiot subjects, but probably not so pronounced. Jimmy cried when he could not solve the problem; Carl, on one or two occasions, stamped his foot and then sat down and sulked. As with apes, emotional outbursts are most likely to appear after the subject has been thwarted in an attempted solution. However, it may be added that one subject was eliminated from the group because of temperamental apathy and inactivity in the experimental situation, and another because of emotional outbursts as soon as he was left alone in the room.

The idiot children were slightly retarded physically, and the group ranged in life age from 8 to 11 years. The chimpanzee subjects, in the investigations referred to, have been mature, or at least in the adolescent stages of growth at the time of experimentation.

The preschool child of this same general level has lived only two to three years, at least one of which has been spent in almost helpless infancy with respect to walking and the other activities which increase the field of experience. Thus, although the ape has reached his prime in motor development, and the idiot has a relatively high development in this respect for his mental level, the normal preschool child is only just overcoming the uncertainties of balance, has small motor equipment and undeveloped manual dexterity. We might therefore suspect that physically the ape is superior, and the idiot child is more capable of competing with him than is the preschool child.

In general, a comparison of the three groups leads us to the conclusion that the behavior of the idiot is very similar to that of both the preschool child and the ape. Any differences noted seem to be toward the primitive type of performance, which may be attributed to the fact that in length of experience and physical development the idiot can compete with the ape more favorably than can the preschool child.

SUMMARY AND CONCLUSIONS

1. This investigation is one of a series designed to investigate certain theoretical aspects as to the nature of feeble-mindedness and to apply the newer methods of psychological experimentation to low-grade feeble-minded subjects. The specific purposes of this study were: (*a*) to confirm some of the tentative conclusions concerning problem-solving among idiots, brought about by a former study of box stacking, and (*b*) to make further comparisons with the performances of this group and those of preschool children and apes. These comparisons were made by means of the investigations of Alpert and Köhler, respectively.

2. The technique of this study was similar to that of Köhler, with several relatively important modifications, namely, (*a*) the subject was unaware of the proximity and observation of the experimenter, (*b*) several lures were employed, (*c*) the lure was inside the pen and the subject outside. Problems were presented which required that the subject, on his own initiative, use sticks found in various positions as the means of obtaining the lure.

3. The subjects of the study were seven feeble-minded boys of idiot level, ranging in life age from 8 to 12 years and in Kuhlmann-Binet mental age from 9 to 38 months.

4. The following conclusions are drawn from the results of the investigation:

a. Adaptive behavior among idiot children may be studied advantageously by the problem-solving technique originally used with apes.

b. As in a former study of box stacking,

1) a greater drop in time occurs between the first and second than between the second and third trials.

2) The ability to solve is, to some degree, correlated with the general level of ability as measured by developmental scales, and

3) Among children of the same level, individual differences play an important rôle.

c. Changes in incentive are accompanied by differences in result. Broadly speaking, any change in incentive, which is accompanied by increased activity, facilitates problem solving.

d. These conclusions, aside from contributing to a knowledge of idiot mentality, reveal the fact that in this group we have material for experimental investigation of the low levels of human intelligence, where novel modes of behavior are aroused which have not been influenced by experience or by inhibitions due to training.

5. The results of the present investigation are compared (*a*) with similar results with apes (as in the previous box-stacking study); (*b*) with normal infants (as in the study of genetic development); and (*c*) with both apes and normal infants in the use of implements. Although modifications in technique prohibit a direct comparison, we may conclude that: (*a*) there are marked similarities in the mode of attack and type of solution employed by the three groups, and (*b*) those differences and gradations noted indicate that the behavior of idiot children in such situations lies between that of apes and that of normal infants.

REFERENCES

1. ALDRICH, C. G. Incentive as a factor in problem solving. *Tr. School. Bull.*, 1931, 27, 121-126.
2. ALDRICH, C. G., & DOLL, E. A. Comparative intelligence of idiots and normal infants. *J. Genet. Psychol.*, 1931, 39, 227-257.
3. ———. Problem solving among idiots. *J. Comp. Psychol.*, 1931.
4. ALPERT, A. The solving of problem situations by preschool children. *Teach. Coll. Contrib. Educ.*, 1928, No. 323. Pp. 69.
5. BINGHAM, H. C. Chimpanzee translocation by means of boxes. *Comp. Psychol., Monog.*, 1929, 5. Pp. 91.

6. BRAINARD, P. P. The mentality of a child compared with that of apes. *J. Genet. Psychol.*, 1930, **37**, 268-293.
7. KÖHLER, W. Mentality of apes. (Trans. by E. Winter.) New York: Harcourt, Brace, 1925. Pp. 336.

*The Training School
Vineland, New Jersey*

LA RÉOLUTION DES PROBLÈMES CHEZ LES IDIOTS: L'EMPLOI DES OUTILS

(Résumé)

Celle-ci est une étude de l'emploi des outils dans la résolution des problèmes, avec des idiots comme sujets, employant une technique semblable à celle rapportée par Köhler dans "La Mentalité des Singes Supérieurs."¹ On a présenté des problèmes qui ont fait au sujet, sur sa propre initiative, utiliser un bâton ou plus en obtenant un leurre au centre d'une clôture carrée de huit pieds. Les sujets ont été garçons arriérés du niveau des idiots, ayant un âge chronologique de huit à douze ans et un âge mental de dix-neuf à trente-huit mois. On a comparé les résultats de l'étude avec les résultats de l'étude des singes par Köhler et l'étude semblable des enfants préscolaires par Alpert.²

L'étude donne les conclusions suivantes: (1) on peut sujeter d'une façon satisfaisante le comportement des enfants idiots à une technique employée premièrement pour les singes; (2) les observations de l'apprentissage chez les idiots dans ces situations suggèrent (a) qu'un plus grand abaissement de la courbe de temps se trouve entre la première épreuve et la deuxième qu'entre la deuxième et la troisième, (b) que la capacité de résoudre est à quelque degré corrélée avec le niveau général de capacité, mais (c) que chez les enfants du même niveau de capacité, les différences individuelles jouent un rôle important; (3) en général, un changement quelconque de stimulant, accompagné d'une plus grande activité, facilite la résolution des problèmes; (4) les sujets idiots présentent un nouveau champ pour l'étude de que les différences et les gradations notées indiquent que le comportement non influencées par l'expérience et l'entraînement; (5) la comparaison de ces résultats avec ceux de Köhler et d'Alpert suggère (a) qu'il y a des similarités frappantes dans la méthode de l'attaque et le type de résolution employés par les trois groupes (idiots, enfants normaux, et singes) et (b) que les différences et les gradations notées indiquent que le comportement des enfants idiots dans de telles situations est entre celui des singes et celui des enfants normaux.

ALDRICH ET DOLL

¹Köhler, W. *The mentality of apes*. (Traduit par Ella Winters.) New York: Harcourt, Brace, 1925. 336 pages.

²Alpert, A. *The solving of problem situations by preschool children*. New York: Harcourt, Brace, 1928. 69 pages.

DAS LÖSEN VON AUFGABEN BEI IDIOTEN: DER GEBRAUCH VON WERKZEUGEN

(Referat)

Die Verfasser untersuchten den Gebrauch von Werkzeugen bei Lösung von Aufgaben indem ihnen Idioten als Versuchspersonen dienten un ihre Methode der von Köhler in "The Mentality of Apes"¹ geschilderten glich. Es wurden Aufgaben dargeboten welche verlangten, dass die Versuchsperson aus eigener Initiative von einem Stock oder mehreren Stöcken Gebrauch machen sollte, um aus dem Zentrum eines acht-Fuss-langen Vierecks einen Lockgegenstand ("lure") zu erlangen. Als Versuchspersonen dienten sieben schwachsinnige Knaben idiotischen Niveaus deren kronologisches Alter sich zwischen acht und zwölf Jahren und deren geistiges Alter sich zwischen neunzehn und acht-und-dreissig Monaten erstreckte. Die Ergebnisse der Untersuchung wurden mit denen von Köhler's Untersuchungen an Affen und Alpert's² ähnlicher Untersuchung an vorschulpflichtigen Kindern verglichen.

Die Untersuchung liefert folgende Resultate: (1) Anpassungsbenehmen ("adaptive behavior") idiotischer Kinder kann befriedigend einem Verfahren unterworfen werden, welches ursprünglich an Affen gebraucht worden war; (2) Beobachtungen über des Lernen bei Idioten in diesen Lagen scheinen anzudeuten, dass (a) zwischen dem ersten und dem zweiten Versuch ein grösseres Ersparniss an Zeitgebrauch stattfindet wie zwischen dem zweiten und dem dritten Versuch; (b) die Fähigkeit, Aufgaben zu lösen, wechselt einigermassen mit der allgemeinen geistigen Fähigkeit, aber (c) bei Kindern mit gleicher geistiger Fähigkeit spielen individuelle Unterscheide eine grosse Rolle; (3) im allgemeinen wird das Lösen von Aufgaben erleichtert durch irgend eine Veränderung des Anreizes, von beschleunigter Tätigkeit begleitet; (4) idiotische Versuchspersonen liefern ein neues Feld für Untersuchungen über menschliche Intelligenz,—ein Feld, worin neue Benehmensweise, nicht durch Erfahrung und Dressierung beeinflusst, hervorgerufen werden. (5) Vergleichung dieser Ergebnisse mit den von Köhler und von Alpert erzielten scheinen anzudeuten, (a) dass die Angriffsweisen und Lösungsformen der drei Gruppen (Idioten, Kinder, und Affen) auffallende ähnlichkeiten haben, und (b) dass die beobachteten Unterschiede und Zwischenstufen anzeigen, dass das Benehmen idiotischer Kinder in solchen Lagen eine mittlere Stellung zwischen der der Affen und der der normalen Kinder annimmt.

ALDRICH UND DOLL

¹Köhler, W. The mentality of apes. (Übersetzung von Ella Winters.) New York: Harcourt, Brace, 1925. 336 Seiten.

²Alpert, A. The solving of problem situations by preschool children New York, 1928. 69 Seiten.

WRITING MOVEMENTS AS INDICATIONS OF THE WRITER'S SOCIAL BEHAVIOR*

ROBERT SAUDEK

THESIS

No writer can at the same time be fully concentrated on both the subject-matter he is writing about and his chirography. In most cases the attention is divided between the two, and the way the writer allows the one or the other part of his task to absorb his attention gives a clue to some traits of his character. But there are also cases of two different extreme types, e.g., those whose attention is almost completely concentrated on the subject-matter, and those whose only concern is to write as beautifully as possible.

The more the writer is concentrated on the subject-matter, the more "natural" is his writing (the term "natural" standing for unconscious, mechanized movements).

The slower a man writes the easier can he deliberately shape his script. A slow writing need not necessarily be "unnatural," i.e., deliberate and artificial, but it *might* be so.

The present paper gives the indications from which we can reliably establish whether a handwriting was produced spontaneously, with quick, unconscious, and mechanized movements, or rather with deliberate concentration on the chirography. These indications, which give us reliable clues to definite traits of the writer's social behavior, have been shown and isolated by a series of experiments and measurements.

HISTORICAL SURVEY OF LITERATURE

Amateurish graphology was known for more than a hundred years. Among the dilettanti who practiced it were men like Disraeli, E. A. Poe, Robert Browning, Sir Walter Scott, etc. Most of their observations about the causal nexus between character and handwriting have been conclusively refuted by tests and experiments of modern graphologists, but they have proved to be right in their

*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office, October 27, 1930.

assumption that social behavior shows in some way either in careless scrawl or in careful calligraphy.

Michon (15) was the first to publish several textbooks based on his rather extensive observations. He claimed that certain deviations from the average style of writing (which he called "signs") always occur together with definite traits of the writer's character. To Michon an average writing was a fully matured writing in the style of his contemporaries. It did not occur to him that there cannot possibly be such a thing as an *average* mature writing when different writing movements were taught in different schools.

Crépieux-Jamin (3, 4, 5) tested all "signs" of Michon, accepted only part of them, and started a new method of psychodiagnostics. At that time psychology was rather a subject of intuition than of experimental tests and all theories were based on introspection instead of on methodical research. Crépieux-Jamin divided all mental qualities into two groups, e.g., the primary and the complex qualities. Only the primary qualities, sensitiveness, activity, ambition, show, according to Crépieux-Jamin, in handwriting, whereas the complex qualities, e.g., jealousy, are the result of a number of primary qualities, and can be found only by psychological considerations. This is the leading principle of his theory of "résultantes."

The German physiologist (of English extraction), Preyer (16), made the rather vague endeavor to test the observations of his predecessors and tried a sort of physio-psychological explanation of individual traits in handwriting. He was a poor experimenter. His experiments were only concerned with the question how far the limbs with which one writes are responsible for the shape of one's chirography. He proved that any writing, regardless of whether it was produced with the right or left hand, with the right or left foot, or with the mouth or any other part of the body, still shows, after some training of the unaccustomed movements, so much similarity to the same man's previous normal writing, that both must needs originate from the cortex. But in all his experiments he stopped half-way. When, to mention one instance out of many, he compared writing produced in daylight with writing done in the dark, it did not occur to him to compare also the writing of blinded people with the writing they produced when still enjoying their eyesight.

Meyer (14), a Berlin alienist, first discovered the fundamental laws on which modern graphology is based. He undertook rather extensive methodical experiments with a great number of persons to

whom he put the task of disguising their handwriting in two ways, e.g., first according to their own ideas and then according to definite rules laid down for this experiment. His experiments, as far as they went, yielded results whose validity has since been confirmed by tests made with quite different methods. Meyer established three definite laws:

1. When attempting to produce a handwriting quite different from one's own, one disguises mainly, if not exclusively, the conspicuous parts of the writing and leaves those features unaltered which do not strike the eye.

2. The longer the experiment lasts, the more the writer is apt to fall back into his normal style.

3. The incapacity to concentrate on several characteristics of the chirography when attempting to write in an alien style shows not only towards the end of a longer text, but even within words of more than three syllables.

Meyer's three laws, though correct, are incomplete. Meyer was not aware of the existence of concomitant factors, and has, therefore, not attempted to isolate their effect.

We know now that time is the main factor which determines success or failure in the attempt to disguise one's writing. Anyone knows from his daily experience that his writing becomes neater, more legible, better spaced, more distinct—in short, "nicer"—when he takes time and trouble to mind his calligraphy. Any experiment to disguise one's hand would show that such an attempt (regardless of whether it is successful or not) retards the tempo of the writing. It takes us less time to write a given number of words in our own habitual style than in an alien style, though, of course, the greater the writer's skill, the smaller his loss of time will be. The optimal speed of the writer can never be attained in a disguised hand.

Klages (10) was not interested in the experimental part of the problem. He accepted practically all findings of Meyer and worked the results of all his predecessors into a system. The merits and shortcomings of Klages' theory are therefore conditioned by those of Meyer's experimental methods, which Klages has made both the absolute basis and the test-stone of his theory.

But his is the merit of having definitely refuted the theory that any single feature in handwriting can bear conclusive evidence of a definite trait of the writer's habitual behavior. If there be any definite relation between traits of character on the one hand and features of

handwriting on the other, such a correlation should needs show in the simultaneous occurrence of a great number of "parallel features," and not only in one single "sign" (parallel features being those deviations from the normal average which originate from one and the same either physical or mental peculiarity of the writer).

He also attempted to develop a new theoretical psychology of character (*Charakterkunde*) exclusively from movements of expression (*Ausdrucksbewegung*), and thus to create a novel method of psychology (*Ausdruckslehre*).

His conceptions are, however, rather vague, not easily intelligible, and based on two main principles, namely, the principle of identity of impulse and expression, and the assumption of an inherited sense of space (*Raumgefühl*) different in every individual. Both assumptions have been refuted.

FIFTEEN FACTORS WHICH COOPERATE IN THE FORMATION OF INDIVIDUAL HANDWRITING

It is only today that we have succeeded in experimentally isolating 15 factors (of which the present author claims to have isolated 11) which come into play in the formation of individual handwriting. In extreme cases a single one of these 15 factors may be so predominant as to paralyze the influence of all the rest. We also know now that it was only when former graphologists were confronted with one of these extreme cases, in which the "character" factor was predominant, that their guesswork proved to be correct.

Since the present paper deals only with one aspect of the problem, we have to restrict ourselves to a general survey only of the 15 factors, and to refer the reader to the literature.

1. *The Writing-Implements.* The writing-implements (pen, ink, pencil, paper, pad) normally do not distort the writing so much as to make its identification impossible to the expert, but there are extreme cases when this might happen. [For more details see (18, pp. 235 ff.).]

2. *The School Copy from Which We First Learned to Write.* Some writing systems differ in the movements from others. Experiments with about 50,000 children who were taught to write in two quite different systems show how far this tuition has influenced their quick writing when they were allowed to choose either system or a combination of both. [For more details see (18, pp. 19 ff.).]

3. *The Degree of Graphic Maturity.* A child writes, first, sep-

arate strokes, each movement being caused by a separate impulse (stroke impulse). In a later stage he writes with letter-impulses, then with word-impulses, and at last with sentence-impulses. Only a fully mature writing shows a rhythmical writing pressure typical of the writer. This has been shown by Meumann (13), and demonstrated and measured by Drever (6).

Any mature writer can arbitrarily reduce the maturity of his handwriting to a lower stage. That is to say a person whose habitual handwriting is excellent can produce an illiterate handwriting; but an illiterate person can *never* produce a perfectly mature writing.

For instance, if we have to discover the writer of an anonymous letter, and if this anonymous letter has been written by a matured-writing person, we can exclude from our list of suspected persons all those whose ordinary handwriting is immature (see 18, pp. 53-55).

4. *Acute and Permanent Physical Disorders.* We are able to determine the physiological condition of the writer only by comparing his handwriting from different periods, although we are as yet unable to make differential diagnoses except in certain extreme cases.

We may distinguish between permanent impediments and those which are only acute. The central nervous system adapts itself, after a time, to the loss of the extremity (the hand) with which we have been in the habit of writing, and also to the loss of any sense (eyesight) which normally plays an important part in the act of writing.

This is the reason why any acute impediment distorts the handwriting far more than a permanent impediment (see 18, pp. 211-218).

5. *The Degree of Writing Routine.* Manual skill in various activities not connected with writing does not necessarily improve the writing routine. People whose professional work asks for precision in the minutest movements sometimes write a clumsy hand, whereas people who are clumsy in any handicraft but write a great deal acquire an almost perfect certainty of aim in the writing movement.

6. *The Speed Actually Applied.* The speed actually applied plays a prominent part in the shaping of any handwriting. We are going to deal with this factor separately.

7. *The Degree of Visual Sensitiveness.* The isolation of this factor was possible only after the Great War, which has provided us with many clinical cases of soldiers who have lost their sight and whose former "sighted" handwriting can be compared with their present "blind" writing (see 18, pp. 216 ff.).

8. *Memory for Forms.* Loss of eyesight makes it impossible to

receive fresh visual impressions. This is why persons who have lost their sight never subsequently modify the shape of their letters. Their visual memory reproduces their former handwriting more and more distinctly, until its shape crystallizes in their imagination into a permanent image. Experiments with sighted writers who wrote either in the dark or blindfolded show different results. The latter can arbitrarily produce as many different styles of the writing as they master when not handicapped by lack of visual control; the first group cannot (see 18, pp. 216 ff.).

9. *Previous and Actual Environments.* Where a strong visual sensitiveness and memory for forms cooperate with the influence of foreign environments, the writer is likely, consciously or unconsciously, to adopt foreign forms quite different from those of his previous writing. Hence the assimilation of the handwriting of alien immigrants to the American style, or of American emigrants to their new environments.

10. *Standard of Education (Intellectual Training and Aesthetic Refinement).* If we compare 300 handwritings of graduated students of medicine, of law, and of science, with 300 handwritings of artisans, privates of the army, and farmers, we shall notice that the 300 educated men write on the average more fluently than the 300 "uneducated." But neither will all 300 educated writers show a scholarly hand, nor all the 300 "uneducated" a clumsy one. About 20 out of each 100 educated writers will write a hand inferior to those of the best 20 "uneducated," whereas 90 of the educated will have written at a quicker rate than the average of the "uneducated." Whether this ratio, experimentally established on the European continent, holds good also for American handwritings, I do not know. Probably Factor 5 will play a certain part, since in America the type-writer is much more in use than in Europe (2).

11. *General Intelligence.* Intelligent people show the propensity to write with sentence-impulse even when their standard of education would suggest a lower degree of graphic maturity, e.g., word-, letter-, or stroke-impulse. The more intelligent the writer is, the more he would concentrate rather on the subject-matter than on the chirography.

Even partly illiterate but generally intelligent people who do not master the orthography of their own language show a marked self-certainty by writing rather quickly, in spite of their self-invented phonetic spelling.

In the same way, thousands of intelligent but uneducated people, at a time when there was no compulsory education, showed greater lucidity of thought and a better general efficiency than hundreds of their unintelligent but wealthy contemporaries who could afford schooling.

Education to a standard above the capacity of the pupil often causes self-consciousness and a feeling of inferiority which retards the writing movement and makes it uncertain.

Kaufmann (9) established the same phenomenon with school children, among whom the intelligent wrote with greater self-certainty and speed than the less intelligent ones.

12. *Concentration on the Subject-Matter.* With this factor I am going to deal separately.

13. *Concentration on the Calligraphy.* This factor also has to be dealt with separately, being a main part of the present thesis.

14. *Characterological Factors.* Characterological factors, e.g., the degree of self-certainty and self-consciousness in the writer's behavior; sincerity or mendaciousness; perseverance or carelessness; activity or inertia; sense of detail or superficiality; reserve or loquacity; passion or frigidity; stubbornness or sociability; vanity or modesty; ambition or indolence; sensitiveness or callousness, equanimity or irritability, versatility, or pedantry; reliability or dishonesty (see 18, pp. 223-289).

15. *Psychopathological Factors.* If given unseeded manuscripts of sane and insane individuals in a haphazard collection, regardless of whether the handwritings of the psychopathic individuals have been written in the acute stage of the psychosis or in periods of remission, we can tell insanity in 70 cases out of 100.

Out of the 100 manuscripts written during the acute period of a psychosis 90 can be recognized correctly.

Only in extreme cases a *differential diagnosis* from handwriting which would agree with the clinical diagnosis can be made. But we can, in most cases, tell from features of handwriting the following symptoms of mental diseases: feeling of inferiority; grandiose ideas; pseudologica phantastica; exhibitionism; depression; euphoria; emotional instability; hypersuggestibility; flight of ideas (agitation); absent-mindedness; defects of memory; incoherence of train of thought; dementia; stereotypy.

To the clinicist it is clear that these symptoms by themselves form only a clue but not a reliable basis for a differential diagnosis, since various of these symptoms occur in different mental diseases (19-23).

THE GUIDING PRINCIPLES IN THE TUITION OF HANDWRITING

An efficient tuition of handwriting is nowadays directed to three main aims: (*a*) distinct and easily legible script; (*b*) quick writing; (*c*) writing with the easiest mechanized movements.

School tuition considers first the reader, and then the writer. It insists on a distinct and legible script regardless of whether the ideal fulfilment of this task involves loss of time, greater strain on the writer, and a concentration on the chirography rather than on the subject-matter of the text. Or, to put it in another way, the teacher's first endeavor is directed to the improvement of the pupil's social behavior.

Only after the legitimate claim of society (of the reader) has been satisfied, does the teacher feel entitled to consider the interests of the pupil, namely, the training of his movements in a way which involves the least loss of time, the least manual strain, and the training of those mechanized movements which do not need any concentration on the actual chirography and give the mind free play to concentrate mainly, if not exclusively, on the wording of the text.

Fifty years ago or so, school instruction started from opposite principles. The wording of the text (style of composition) was considered first, and, since the simultaneous concentration on the wording and the calligraphy was considered impossible (as, in fact, it is), the pupils were expected, first, to draft their letters or compositions without minding their calligraphy, and only afterwards neatly to copy their own text, then, of course, with calligraphy as the only task in their minds.

Passionate people, when concentrated on the subject-matter only, wrote quite as carelessly at any time as they do nowadays. Some people even suited their style to their actual moods. Queen Elizabeth, for instance, wrote vertical calligraphy (see Figure 1), but fell into a slanting style when in one of her fits (see Figure 2). The English aristocrats up to the nineteenth century used to draft their intimate letters in a careless longhand, minding neither legibility nor orthography, and had these scribbles neatly copied and correctly spelled by their secretaries.

Nowadays we hardly ever keep copies of our long-hand letters and would hate the idea of writing our letters twice.

CAREFUL AND CARELESS WRITING

If speed and concentration on the subject-matter, on the one hand,

and retardation of the writing tempo and concentration on the calligraphy, on the other, were the only two factors to determine the legibility and the neatness of the writing, then any fully matured hand would show us in which way and to what degree the writer compromised between either the easiest self-expression or some social considerations. It is obvious that a self-centered man of quick thought could be so much concentrated on the ideas he is actually expressing with his pen that he would almost completely forget the prospective reader and, with the least possible strain, would produce an almost illegible scribble. And quite in the same way a mentally dull writer, whose only ambition is to impress his reader by the virtues of his

Your loving sister Elizabeth

FIGURE 1

hand, as provided for a heart
that shall ever be filled with
your affection of your
Loving and sincere
sister
Elizabeth

FIGURE 2

calligraphy, would write a "beautiful," pedantically flourished hand, not only not minding, but actually enjoying, his deliberate, slow movements.

But things are not so simple. In most cases additional factors play a part. Instead of *deliberately* controlling his movements, a writer may do so *unwittingly*, because aesthetic care for his graphic expression has become second nature to him. In the same way, a man with real sense of order would hardly allow himself to yield so much to what we call the "sentence-impulse in writing." Because of the love for orderliness, he could not help minding distinct spacing even when writing at his optimal rate of speed or when carried away by some momentous affect which he expresses with his pen.

A comparison of slow and careful with quick and spontaneous writings of the same person shows us how differently he reacts to different environments, how his state of mind varies with the audience he addresses in his writing.

WHY WE MUST FIRST DETERMINE THE ACTUAL SPEED APPLIED IN A WHOLE MANUSCRIPT OR IN PARTS OF IT

In my view, the psychology of handwriting has only begun since I have made it a fundamental point first to establish to what degree a handwriting has been produced spontaneously, with mechanized movements, or, should this be only partly the case, to what degree the writer concentrated on his calligraphy.

From what has already been said, it is clear that to answer this question we must first establish how quickly a manuscript has been written. Any writing produced at the writer's optimal speed is necessarily his habitual writing. Of course it is quite possible that a man's previous habitual writing was an illegible scrawl, that he has devoted a good deal of his time methodically to compensate for this shortcoming, and that his present "beautiful" hand has become second nature to him. But even then his present optimal speed will be inferior to his previous one. Should he try to increase his present speed to his previous standard, his calligraphy would necessarily deteriorate.

RATE AND TEMPO OF WRITING

By training we can acquire a skill in accelerating and retarding our writing tempo, but there is a limit outside which neither training nor skill will carry us. We may assume that this optimal acceleration

and retardation of the writing tempo is in some way typical of the writer's personality. In each individual the increase and the decrease of the speed depends on the speed with which he writes when his mind is fully concentrated on the subject-matter.

Suppose a person is used to write on the average 140 letters in a minute and trains himself to a higher efficiency—he will soon discover that after having attained a speed of, say, 200 characters in a minute he cannot improve this performance any more. Should he insist on doing so, definite indications in his writing would warn him that it is no use trying further. His writing would deteriorate because of his irritation, he would write with superfluous flourishes, thus extending the path of his nib instead of shortening it, he would fall into lapses, e.g., write parts of the words or characters twice (perseveration), and omit others, etc. Quite in the same way, when writing at a gradually reduced speed and concentrating exclusively on his calligraphy, his writing would first gain in neatness, distinctness, etc., but would later also deteriorate and show plain indications to warn him that he is overdoing things. Also, this deliberately slow writer would, because of his irritation, fall into perseverations and omissions, and trace strokes which, instead of being steady, would grow uncertain, tremulous, and broken.

The reader who wants to check the correctness of this statement should first write (with a stop-watch) a test of, say, 15 lines at his habitual rate. If this performance took him a seconds, he should try to write first at the rate of $2a$, and then carefully copy his text at

the rate of $\frac{a}{2}$. He will notice that he was more successful when retarding his writing tempo beneath the habitual normal, but that he was less successful or failed almost completely when trying to accelerate his habitual tempo up to the double. On the average, he will first notice the deterioration of his slow and careful hand when

writing at a rate of $\frac{a}{6}$.

From this experiment he may safely assume that when writing with a sentence-impulse he does so at his optimal rate. This is rather plausible, because any natural writer is likely to keep pace as far as possible with the tempo of his thought, and because we all think quicker than we write. To articulate a thought takes, say, a second; to write its wording down may take 20 or 100 times as long.

The reader who has made this experiment should keep the three specimens of his writing at hand to check in his own writing the correctness of the following laws of the writing movement and indications of the writing tempo, on which modern graphology is based. (A similar experiment is illustrated in Figure 7 *a, b, c.*)

THE LAWS WHICH RULE THE WRITING MOVEMENT

1. Whenever pen is put to paper the nib first pauses before starting the actual movement. It takes a certain time to find one's way about on the paper, and this time can be measured. In slow-motion camera pictures of the writing movement, the first and the second exposures are always identical, which proves that during the time in which these first two exposures were made the nib was at a standstill. Freeman (7) was the first to make these slow-motion photographs at a controlled speed of 25 exposures in the second. If we accept his measurements as reliable, we may assume that the shortest initial adjustment of the nib takes $2/25$ of a second. Where the necessary loss of time is reduced to this minimum, we may assume that the writing movement was quick and spontaneous, and that the attention of the writer was neither distracted to his calligraphy nor retarded by any doubts about wording and spelling.

The slightest doubt in the writer's mind about wording, spelling, and shaping of the script retards the writing movement from the very start. The writer records these mental impediments on paper in tracing any retardation with his pen. The shape of these traces (initial adjustments) tells us an eloquent tale. Not only can we read from the shape of the initial adjustments what loss of time they involved, but also to what mechanical, mental, or physical cause they were due. There are, of course, various possibilities. The start of the writing movement may be elaborate, circumstantial and retarded:

a. because the writer is temporarily or permanently impeded by lack of coordination in his musculature;

b. because he has some vague doubts as to the orthography of a certain word;

c. because he ponders, even if only for part of a second, about the wording of his text;

d. because he lacks skill and writes longhand notes only rarely;

e. because his mentality is dominated by self-consciousness, fear, indecision; or

f. because he has to overcome some subconscious reluctance as-

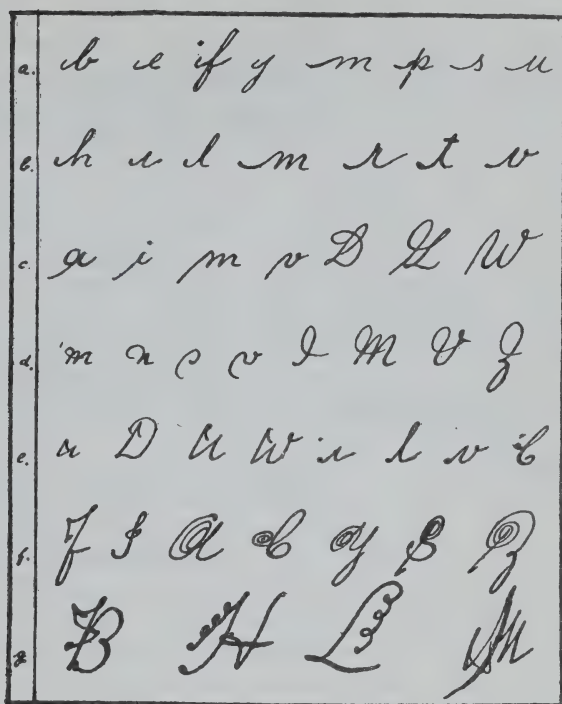


FIGURE 3

sociated, in the psychoanalytical sense of the term, with the word which he starts to put on paper.

From what has been said it is evident that we need additional symptoms to reduce the number of possibilities in such a way that only one definite cause for the retardation of initial movement can be assumed.

I propose to give here only a few instances just to show the method applied, without, however, claiming that these instances exhaust all possibilities. For a more thorough explanation of the subject I must refer my readers to my books.

Figure 3 shows instances of various shapes of initial adjustments. The first line shows the easiest, which take the shortest space of time; as we proceed down the column the adjustments become progressively slower, heavier, and more elaborate.

Berlin 18 July 1879

Dear Sir!

Allow me to send you the conditions
attached to the prize offered by the
Emperor of Germany for the best work
on Diphtheria.

Yours most respectfully

B. von Kagerbauer

FIGURE 4

worried over such a thing
and show myself how
there was to upset me
I have learned how we
I was to have borrowed

FIGURE 5

To find out whether lack of coordination in the musculature of the writer is responsible for marked initial adjustments in his writing, and, if so, which part of the limb is in an impeded state, we must keep in mind the results of those experiments by which it has been established which rôle each part of the limb plays in the writing movement. I propose to give here only the plainest of these results, namely, those necessary for the elucidation of our present problem.

At a normal holding of the pen, when the penholder rests between thumb and middle finger, whereas the first finger rests on the penholder, the thumb slightly presses against the holder in the direction from left to right, and pushes the pen forward along the line. In this way the thumb is responsible for the short sideward movements between the downstrokes of small characters like *m*, *n*, *u*.

Somewhat longer distances along the line, e.g., the sideward movement within short words of no more than two syllables, are produced by the slight rotation of the wrist.

The bigger distances along the line from the beginning to the end are covered by the movement of the forearm.

If now the thumb is unable to go on with this mechanized movement by which the pen is pushed forward when a character like *m* or *n* is written, whereas the first finger has not lost the capacity of its mechanized movement, this incoordination must necessarily show in the handwriting; the normal distance between the downstrokes of the character *n* is lacking, and the two downstrokes which should be parallel *near* each other *cover* each other.

Figure 4 shows this phenomenon plainly in all *m*, *n*, and *u*'s throughout the text; the *n* in the first word, *Berlin*, is apparently written with one downstroke, but in fact with two strokes of which the second covers the first. We see the same in the *m* in the second word of the text *me*, where the third downstroke covers the second; and again two words further on, in the *n* of the word *send*, where the second downstroke covers the first, and so on. Furthermore, we see that all strokes are broken and atactic. From these two concomitants we know that the marked initial adjustments of Figure 4 are due to deficient musculature.

Figure 5 also shows marked initial adjustments, but even the layman can see that they are not due to deficient musculature but rather to lack of both writing routine and skill.

2. Everyone writes in such a way that the speed of the moving pen is constantly varying. This variation of speed depends on, and

is adapted to, the form of the letter which is being written. Rounded letters or parts of letters are produced without a pause in the transition from one direction to the other (the speed is indeed *always* lessened, but the pen is *never* brought to a standstill), but *no one* can produce an angular formation without stopping for a fraction of a second before the transition from one direction to the other (7, 18).

Since tremulous strokes, by reason of their nature, can only be written with a number of changes of direction, while, on the other hand, every "broken" form necessitates turning a corner, it follows that no one when writing rapidly can produce tremulous or broken strokes (18).

3. Everyone writes long strokes in a quicker tempo than short strokes. This is explained by the fact that the movement of the pen over the paper is accomplished more rapidly, for example, when we are writing the downstroke of the letter *f* than when we are writing the downstroke of the letter *i*; for instance, if in our writing the downstroke of the *f* happens to be five times as long as that of the *i*, we do not take anything like five times as long to write it, but less; and if we compare the shapes of the three letters *a*, *d*, *q*, it is quite evident that the path followed by the point of the pen in writing the letters *d* and *q* is longer than that in writing the letter *a*—that is, in *d* it is longer by the upper portion of the final downstroke, and in *q* by the lower portion of the final downstroke. Nevertheless, when writing spontaneously—that is, quickly, and without either physical or technical impediments—everyone writes all three letters in precisely the same time. The speed of the pen is distributed, so to speak, over its whole path; this is necessary to the formation of the letters as units. The acceleration in writing *d* and *q* begins during the writing of the oval, and not only in the last downstroke (7).

4. Everyone writes in such a way that he produces a simple straight stroke at varying speeds. The speed of the movement increases to a maximum and decreases again towards the end of the stroke (1, 7).

5. The laws of inertia will prevent any writer from bringing the movement of his pen to a standstill in accordance with his intention of the moment. If we apply this doctrine to the simplest of all graphic formations, namely, the dot, it means that no one when writing at a high rate of speed is capable of making a dot which actually has the form of a full stop, but that he will, instead of this, produce a form which, when examined through the microscope, and often with the naked eye, has the form of a comma or accent (14).

6. In all European and American writing systems (as distinct from Asiatic) the writing movement progresses from left to right. In addition to this general movement to the right, subsidiary movements in many other directions are required for the correct formation of the letters. In quick writing the rightward-tending movements are performed with greater vigor and emphasis, while those tending in other directions are correspondingly neglected. This is shown in the strokes committed to paper: those running to the right are bold to the point of exaggeration, while the rest are stunted and incomplete. Consequently, the letters are not properly formed.

Every letter assumes a characteristic shape when the writer is driving his pen at full speed towards the right. Take the letter *g*. After forming the oval at the top, we were taught to make a long downstroke and then, by turning the pen towards the left, to produce a loop. That is, while we are hurrying towards the right, we are asked to make a stroke towards the left. In fact, our general urge overcomes the dictates of the copybook, and we omit the left-

I	II	III
o d v w	o d v w	o d v w
i m n u	i m n u	i m n u
r s x z	r s x z	r s x z
y j k f	y j k f	y j k f
l g t j	l g t j	l g t j
c h k t	c h k t	c h k t
p R D K	p R D K	p R D K
b E H L	b E H L	b E H L
s o a g	s o a g	s o a g
n u v w	n u v w	n u v w
x z z	x z z	x z z

FIGURE 6

ward loop altogether, either leaving the letter unfinished with a straight downstroke, or adding a rightward upstroke which transforms the *g* into a *q*.

Column I of Figure 6 shows the forms of certain letters as prescribed by the copybook, and Column II the forms they tend to assume in rapid writing, owing to the general tendency to the right.

The same tendency also affects our arrangement of space. The general body of writing shows a drift towards the right. The left-hand margin grows wider and wider as we proceed down the paper, because we begin each line a little nearer to the goal we are striving to reach, viz., the right-hand end of the line.

Further, a vigorous urge to the right causes the whole script to become wider. This is shown particularly by the small letters *m*, *n*, and *u*. Every man and woman writes with a definite, individual rhythm, which dictates, among other things, that a certain interval of time, characteristic for each person, occurs between the writing of the separate downstrokes of these letters. During this interval, which measures only a fraction of a second, either a slight turn of the joint or a faint pressure of the thumb brings the finger further to the right. It is obvious that when writing quickly we shall cover a greater distance in this given time than when writing slowly. Consequently, these three letters, *m*, *n*, and *u*, acquire a narrow form in a slow writing and a wide form in a quick writing.

As all rapid, spontaneous writing tends to the right, so when the pace is retarded and the writer's spontaneity checked, a tendency to the left appears. This shows itself in a decreasing left-hand margin and in an exaggeration of the leftward movements required for the formation of the letters. Examples of the distortion of letter-forms due to a tendency to the left are shown in Column III of Figure 6.

While a tendency to the right is a natural result of letting ourselves go, a tendency to the left is a proof of the contrary condition, i.e., the curbing of this spontaneity. The more we restrain ourselves the narrower our writing grows, and this is again clearly shown in the forms of the significant letters *m*, *n*, and *u*. In extreme cases our pen may make no forward movement at all, so that the distance covered in the interval between forming the downstrokes of these letters is reduced to nil. Then, instead of being parallel with each other, the strokes coincide, forming what is called a *covering stroke*. The covering stroke proves that at this point the rightward movement had completely ceased (18).

7. No one, when writing rapidly, can make purposive movements with complete spatial certainty. To take a simple example, in rapid, spontaneous, automatic writing no one can place the dot exactly over the *i*. According to the direction which his pen is following at the moment, the writer will place it too far to the right or left or too high up (18).

8. The movements of backhand writing require 30 per cent more time than the movements of slanting hand (8, 12).

9. We write *downstrokes*, i.e., movements directed towards the body, at a quicker rate than those in the opposite direction (8, 12).

10. The movement is retarded the moment the writing pressure is increased (18).

11. Any mechanical impediment (for example, a paper fiber in the pen) will substantially retard the speed of writing (18, 24).

Any single part of the writing is a record of the writing movement unwittingly traced by the writer. The shape of these traces bears evidence of the intensity, acceleration, or retardation of this movement, telling us at the same time whether the retardation was due to mechanical or physical or rather mental inhibitions, or again, whether the quicker tempo in some parts of the writing was due to some sudden mental impulse.

Although quickness or slowness of the movement is by itself an important trait of character, we are, as a matter of fact, more interested in the relative speed of the writing movement than in the determination of its absolute rate. Or, to put it in another way, it does not interest us so much to establish how many characters a person has written in a given time, but rather whether he kept up an approximately equal tempo or whether he wrote parts of his text at a slower rate than the rest. The knowledge of the laws of the writing movements and of their effect on the shape of each part of the writing has enabled me to draw up a table of indications of slow and quick writing movements from which we may establish the relative speed at which the various parts of a writing have been produced.

Figure 7*a* has been written in 11 3/5 seconds at the writer's habitual optimal speed, 7*b* at the maximal rate of 8 1/5 seconds, whereas 7*c* has been written very carefully indeed, and took the writer 82 seconds. If now we check Table 1 on these short specimens, we shall find as many indications of speed in the quick writing and as many of retardation in the slow writing as so short a text

can possibly contain. In the slowest writing (Figure 7c) we can observe with the magnifying glass even a slight tremor (S1), especially in the ovals; the *i*-dots are placed very accurately (S3) and are really shaped like dots and not like accents, and the *t* is crossed on the right spot; the letters are shaped very carefully (S4); the style is uniform and regular (S6); some characters show initial adjustments, e.g., the *j* in *jumps*, and are shaped in a narrow style (S8).

If now we analyse the tempo of 7a, written at the writer's optimal speed, we see that the strokes are fluent (Q1); that the diacritic signs are very inaccurate, and that instead of *i*-dots there are accents (Q3); that the bar does not cross the *t*, but starts from the right side of the character and is prolonged far to the right (Q7); that

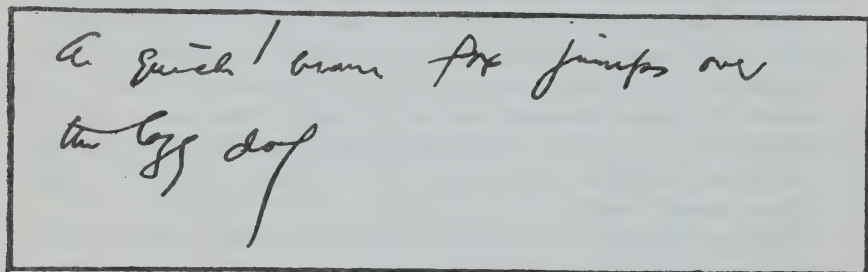


FIGURE 7a

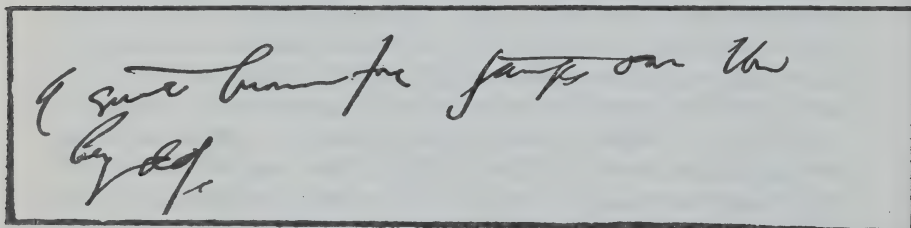


FIGURE 7b

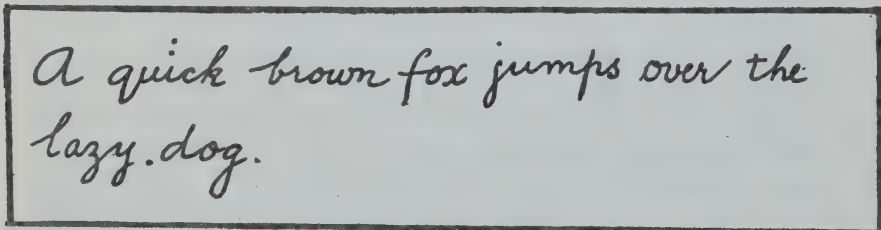


FIGURE 7c

TABLE 1

INDICATIONS OF THE SLOWNESS OR QUICKNESS OF THE WRITING MOVEMENT

Slow (<i>S</i>)	Quick (<i>Q</i>)
1. Wavering, bent, and broken strokes.	1. Fluent, unbroken strokes with rhythmic alternation of pressure do not indicate quick writing; they merely show that the muscular system functioned normally.
2. Lines sinking towards the right.	2. Lines rising towards the right, where the rise is not due to an abnormal position of the paper.
3. Accurate placing of <i>t</i> -bars and <i>i</i> -dots.	3. Inaccurate aim in placing <i>t</i> -bars and <i>i</i> -dots, and production of commas or accents in place of dots.
4. Careful formation of letters, especially towards the end of words.	4. Letters clipped and incomplete, becoming almost illegible towards the ends of the words.
5. Signs of frequent adjustments, either final or within the words (in both cases revealing a pause in the writing movement); letters connected by angles instead of curves; superfluous dots (resting-points of the pen on the paper); marked pressure in parts of downstrokes only instead of in the whole downstroke; frequent change of pen-hold.	5. No readjustments within the words; no final adjustments; connection of <i>t</i> -bars or <i>i</i> -dots with succeeding letters; linking up of words; linking up of figures in a number.
6. Uniform, regular style of writing in which neither capitals nor medium-sized letters are in any way emphasized and the writing is kept at a constant angle, with regular spacing and a general approach to the school copy.	6. Irregular writing with fanciful ornamentation of letters and frequent changes of style; some letters are written with a marked slant, others are more upright, or are quite vertical or even sloped backwards.
7. Frequent signs of tendency to the left, shown by decreasing left-hand margin and characteristic formation of letters (Figure 6, Column III).	7. Frequent signs of tendency to the right, shown by increasing left-hand margin and characteristic formation of letters (Figure 6, Column II).
8. Heavy, elaborate initial adjustments; narrow space between the downstrokes of <i>m</i> , <i>n</i> , and <i>u</i> ; reduced space between letters within the words; covering strokes.	8. No initial adjustments; wide space between downstrokes of <i>m</i> , <i>n</i> , and <i>u</i> ; increased space between letters within the words.

the letters are clipped and incomplete, becoming almost illegible towards the end (*Q4*); that there are not only no readjustments within the words but that the *t*-bar is connected with the succeeding word (*Q5*); that the writing is irregular (*Q6*); that there are frequent signs of the tendency to the right (*Q7*) and no initial adjustments (*Q8*).

Therefore the script written at the optimal rate shows all our indications of speed but two, being neither wide nor rising. Figure 7*b* shows the same man's writing; it is not only much wider and has rising lines, but contains also all the other indications of speed much more markedly.

Let us now test our table on some more lengthy writing. At a first glance the handwriting of Figure 8 looks vigorous, bold, energetic, and quick. But we arrive at a quite different conclusion if we examine its individual features and establish its relative speed from the indications of our table.

The following analysis may serve as an illustration of how to apply our method. We proceed by asking ourselves whether the features given under the various numbers in both columns of our table do or do not occur in the writing.

A rectangular box containing handwritten text in cursive script. The text is arranged in three lines: "I am feeling" on the top line, "fine and have" on the middle line, and "never had a" on the bottom line. The handwriting is bold and somewhat slanted to the right.

FIGURE 8

	We notice:	We note:	
		Quick	Slow
1.	Fluent, unbroken strokes with rhythmic alternation of pressure	1	
2.	The direction of the lines is straight	-	-
3.	The <i>i</i> -dots, in the shape of accents, and the <i>t</i> -bars are placed far ahead of the corresponding character	3	
4.	The letters are written in a bizarre way, but are not incomplete, and sometimes their size increases towards the end of words		4
5.	Frequent adjustments at the end of words; frequent change of pen-holds		5
6.	Irregular writing with bold movements, fanciful ornamentation of letters and frequent changes of slant	6	
7.	Frequent signs of tendency to the left		7
8.	Initial adjustments; covering strokes (in the small letters <i>m</i> , <i>n</i> , <i>u</i>); the connecting upstrokes cross the following downstrokes turning back with the movement towards the left instead of proceeding straight on; this causes a marked retardation of the movement which normally would proceed along the line from left to right		8

We have noted that there are only three indications of speed (1, 3, 6), but four indications of retardation (4, 5, 7, 8), which in our terminology means that this writing is primarily slow, but accelerated by some concomitant causes.

According to Law 3 of the writing movement, we write longer strokes at a greater rate of speed than small strokes. But this increase of the rate does not mean a quicker advance along the lines. It is due to a propensity to big, expressive movements, but the effort does not yield adequate results, since it does not increase the number of characters written in a given time. To the same strong impulse (impatience) are also due the diacritics placed far ahead of the skeleton of the character to which they belong, whereas the remaining indications of speed, e.g., rhythmical alternation of pressure and unbroken strokes, do not bear evidence of mental accelerating causes but only of physical ones, being symptoms of a properly working musculature.

Even in this preliminary stage of our examination we can already read some characteristics of the writer's social behavior. With great effort and overemphasis of her modes of expression, the writer of

Figure 8 shows a rather low efficiency in her actual performances. Her movements are retarded because she concentrates more on her calligraphy than on the subject-matter. We may even assume that she pays for this satisfaction of her urge for ostentatious behavior by her inefficiency. She is interested more in the impression her work makes on others than in the actual virtues of her performances.

The reader will rightly object that we base too far-reaching conclusions on small circumstantial evidence. How do we know that the writer is concerned with her calligraphy and not impeded by some other inhibition in the quick execution of her writing? It is true that we can discard the interference of some physical impediments because we have found that the muscular tonus of the writer was quite all right. But can we not assume that the writer's impulse to write quickly has been impeded by some mechanical obstacle, by the lack of skill, or, say, by slow train of thought which made her stop after every word?

To establish definitely which of the various possible causes was responsible for the retardation of movements, we do really need further evidence. If only the care for calligraphy prevented the writer from quicker writing, and if no other concomitant cause could possibly have cooperated in her case, other indications must be present in her writing.

THE SYMPTOMS OF "NATURAL" AND "UNNATURAL" WRITING

The disguising of one's own handwriting involves great concentration on the shaping of the script. We may, therefore, assume that methodical experiments with disguising of the habitual style of writing will reveal those features which are typical of an increased concentration on the act of writing and possibly show also the limits beyond which a writer is unable to concentrate on the experimental task of disguise.

We know that already 30 years ago Meyer (14) had established that some definite features of one's habitual hand can be more easily disguised than others. But Meyer's findings were confined to the German script only, and, as has been proved experimentally (18), do not hold good for other systems of writing. There is only one rule which holds good for all handwritings, European and Asiatic, namely, that, within his own limits, the disguiser of his writing will be the more successful the more time he is allowed for his task. Otherwise, the degree of difficulty depends mainly on three factors,

e.g., (*a*) whether the deliberate style is or is not consistent with movements typical of the writer's habitual style and with those writing movements in which he has been trained when first learning to write at school; (*b*) whether the shapes to be disguised form the conspicuous parts of the writing or such parts as do not strike the eye; and (*c*) whether the whole system of the writing has to be altered, or some peculiarity only, not connected with the general style of the writing, is to be disguised and therefore must be permanently kept in mind by the writer.

It has been established by numerous experiments made at varying rates of speed that these three factors cooperate in such a way as to allow the drawing up of a definite scale in which each item shows us the degree of the difficulties which stand in the way of the writer who tries deliberately to produce this particular feature in his writing.

SCALE OF THE DEGREE OF DIFFICULTY ENCOUNTERED BY
AMERICAN¹ WRITERS (NUMBER 1 BEING THE LEAST,
NUMBER 15 THE MOST DIFFICULT)

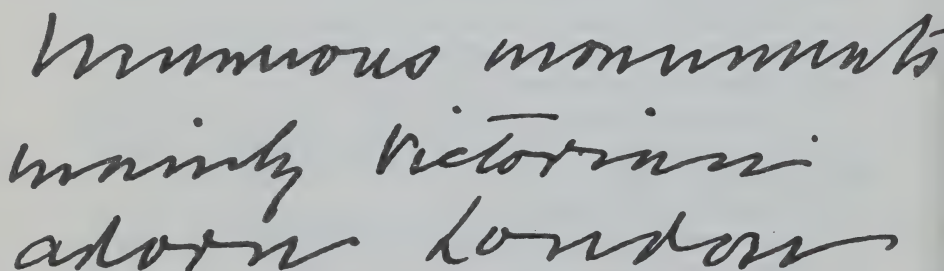
1. Retardation of the usual speed of writing.
2. Enlargement or reduction of the handwriting as a whole.
3. Alteration of the angle of writing (vertical instead of slanting and vice versa).
4. Consistent alteration of pressure throughout the manuscript.
5. Striking letter-formations and distribution of space.
6. Employment of other styles of writing.
7. Accentuated lower projections.
8. Interruptions.
9. Width instead of narrowness and vice versa.
10. Angularity.
11. Arcades instead of garlands or vice versa.
12. Simultaneously accentuated upper and lower projections.
13. Threadlike connections.
14. Accentuated upper projections.
15. Inconspicuous letter-formations and inconspicuous peculiarities of spacing.
16. Undisciplined rhythm.
17. Great acceleration of the habitual speed of writing.

¹The word "American" is here applied to those writers who had their first instruction in American schools, have permanently lived in American environments (U. S. A.) and have not kept on extensive relations with foreign long-hand correspondents.

As an instance of how this scale holds good for American writings, and not, say, for German, we may mention that in all German schools two styles are taught simultaneously, e.g., the "Roman" rounded style and the angular German style. Therefore, hardly any German would find it difficult to write either style. This is the reason why, in the German scale, angular style would not stand in place 10, as it does in the American, but rather in place 2.

Figure 9 shows how a writer who normally writes in a rounded style tried to write an angular hand. At the beginning the writer said he found his task difficult, and began to write very slowly. That he did encounter considerable difficulty is shown by the broken downstroke which begins the second letter *u* in the first word *Numerous*. In the third letter, *m*, the top connections of the second and third strokes are not angles at all, but distinct curves. In the next word the writer succeeded better, but only at the expense of speed. Then he was asked to continue more quickly, and again the angles lost their sharpness. At the beginning of the third line, he made another attempt to achieve his purpose, and the strokes became larger, but still the angles were not sharp. The more trouble he took, the less he succeeded, for the last word, *London*, again very large, contains a broken stroke.

This single experiment may serve as an instance for others to illustrate how difficult it is to disguise a *single* writing-symptom which is placed in the lower half of our scale (24). We notice that, only when taking, say, at least four times the time of his normal writing, the writer succeeds in producing clumsy angular forms for a short while, but falls back immediately into the rounded style as soon as he tries, and were it only slightly, to increase his speed.



Numerous monuments
mainly Victorian
adorn London

FIGURE 9

To understand what the terms *conspicuous* and *inconspicuous* features mean, the reader should place before him at a distance of, say, a yard, some full-page manuscript and try to describe its mode of spacing. He will easily notice whether a broad margin of white space was left at the top, at the bottom, and at the left hand of the written matter, but he will not be able to catch at a first glance some slight variations in the left-hand margin, and even less some irregularities in the right-hand margin. Still, a good observer might notice these less conspicuous details, but it takes even an expert some time to determine definite variations in the horizontal interspace of the words within the line, and in the vertical interspaces between each line. And, again, it takes a thorough examination before he can determine in which typical way the writer is accustomed to make quite superfluous interspaces within those words which he happens to write in separate parts. The first-named peculiarities keep the third place in the following scale, the latter, being more difficult, are put down under 12. In the following scale the various features are drawn up in such a way that the most conspicuous characteristics of the writing appear under 1, and the least conspicuous under 15.

SCALE OF THE DEGREE OF CONSPICUOUSNESS OF WRITING FEATURES

1. Largeness or smallness of the writing. In this connection it will, of course, appear that different observers have different ideas as to what is large and what is small handwriting. The layman's ideas will differ from those of the graphologist. But, since it is precisely this experiment which determines a graphological definition of large or small handwriting, the graphological definition can only be given after the final results of the series of experiments. A certain proportion of the observers will

a. regard a handwriting as large if the capitals and medium letters and long letters are greatly enlarged; others when

b. a handwriting is considered to be too large in proportion to the size of the paper, for example, when it is so large that only a few lines can be written on a small sheet of paper. As a result of this false optical orientation, it may happen that such an observer will declare that characters which are absolutely smaller are larger than characters which are absolutely larger, if the latter are written on a larger sheet of paper.

c. *Scarcely one observer in a hundred will take the size of the*

small letters as his standard of judgment. Because of their inconspicuousness the small letters are taken as a standard; their size is normal, i.e., neither big nor small if the down strokes are about $1/10$ ".

2. The degree of emphasis or "shadowing," and whether the strokes are noticeably thick or noticeably thin.

3. The external arrangement of the script—that is, the distinct spacing of the words and lines, or the insufficient distance between the individual words, horizontally, or the individual lines, vertically; the width or narrowness of the upper and lower margins; very marked widening or narrowing of the left- or right-hand margin; individual details, such as the fact that a postscript is tucked away to one side, right at the bottom of the page.

4. Angle of writing; that is, very slanting or vertical or back-hand writing.

5. Marked emphasizing of individual parts of the text, either by underlining, notes of exclamation, the employment of different systems of writing, or strong pressure, conspicuous enlargement of characters, complicated and decorative forms, etc.

These first five impressions of the handwriting received by the observer may be regarded as a sort of placard impression (the impression received at a first glance, even at a distance). They can be noted even before the observer takes the manuscript in his hand for closer examination. Now, only, does he proceed to more detailed observations.

6. The handwriting is careful or negligent.

7. The form of the capitals and the long or medium letters is noted, particularly if their original formation or their adherence to unfamiliar styles of writing produce an impression of strangeness.

These seven points should be dealt with *before* the observer proceeds to *read the manuscript*. Only at this stage of the experiment may the observer proceed to *read* the manuscript, in order that he may be in a position to make further observations.

8. Degree of legibility. In the case of manuscripts which are difficult to read, most observers are unable to give any clear explanation of *why* they are illegible. They commonly speak of the carelessness of the writer, but are not in a position to say whether the carelessness is due to speed or the contrary, namely, to laziness in connection with slow execution [see my table of the equivocal signs of speed (18)]. Neither can they distinguish whether it is due to the deterioration of the letter-formations; and the reason why

they cannot do so is this, that *they are not aware that their reading is partly guesswork*. If they were in a position to do so they would be able to assert that they do not read the stunted word-endings of familiar words, but guess them. Indistinct and imperfectly formed letters in the middle of words in combination with legible word-endings represent an incipient self-control, a reaction on the part of the writer against the indistinctness which is producing a disagreeable impression on him—a reaction which is plainly reflected in his characters.

9. Individual letters representing quite different letters: for instance, *ee* instead of *u*, or *uu* instead of *ear* (*Dear*), etc. But scarcely one untrained observer in a hundred will at this stage of the description be capable of stating that the letter *u* has been written instead of *n* (that is, a garland form), or *n* instead of *u* (that is, an arcade formation). Untrained observers are not capable of doing so because they all guess *u* and *n* correctly from the context, whether they diverge into garland form or arcade.

Most intelligent observers should be able to note all these nine points; but all other points should be reserved for quite particularly good observers.

10. The relation of the upper and lower projections. For example: the upper projections are emphasized while the lower projections are neglected, the lower projections are emphasized while the upper projections are neglected, etc.

11. Width and narrowness.

12. The less noticeable peculiarities of spacing. For example, there is an interval between the words, but it is not sufficiently large; the lower projections frequently invade the line underneath; the left-hand margin grows a *little* wider; there is a larger space between the end of one sentence and the beginning of the next than between the words of a sentence; the lines are closer together towards the end of a page than at the beginning, etc.

13. The peculiarities of the short letters.

14. The manner in which the diacritical signs are placed.

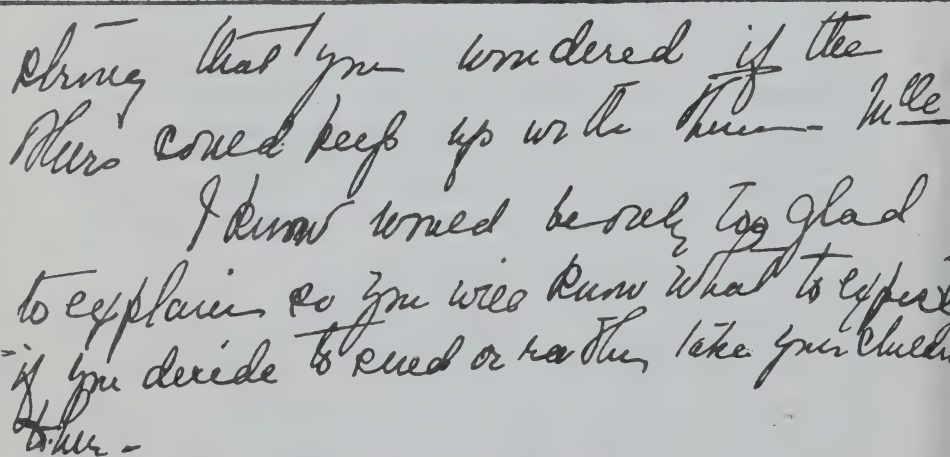
15. The particular kind of hair-stroke by which the letters are connected with one another. It will, however, be found that tremulous forms are noted sooner than thread-connections.

We may assume that handwritings which are written quickly are habitual and natural writings, because, in order to write at the optimal speed, the writer must concentrate on the subject-matter.

But there is also a sort of habitual writing which is, to a certain degree, unnatural in the graphological sense of the word; in such a case the writer has deliberately altered the style of his chirography, and has trained himself in this adopted style sufficiently to produce it with mechanized movements. In daily life we speak of a man who has played a part until it has become second nature to him. He can mislead only those of his correspondents who do not happen to have studied experimental graphology. To the expert his handwriting plainly reveals not only the part he endeavors to play, but also his real nature which he wants to disguise. The style of his writing will appear consistent to any observer who cannot tell the difference between its conspicuous and inconspicuous parts simply because he has never noticed the inconspicuous ones, in the same way as the writer, when first adopting an alien style, adopted only its conspicuous parts.

To the inexperienced reader Figure 10 may appear quite natural and consistent in its style, but an examination on the strength of our present knowledge tells us a different tale.

Let us first examine whether the text was written with what we have called a sentence-impulse, I mean, with full concentration on the subject-matter, or if the attention was rather diverted to the calligraphy.



showing that you wondered if the
 who could keep up with them - in the
 I know would be only too glad
 to explain so you will know what to expect
 if you decide to read or rather take your choice
 to read -

FIGURE 10

We first ask ourselves whether there are numerous initial adjustments. There are indeed. The first word reads *strong*. Without going into a detailed description of the movements with which the *s* was traced on the paper, it is evident from the first covering stroke that a strong retardation has taken place. What has actually happened is this: first the writer wrote a slanting downstroke, then up she went precisely in the opposite direction, covering her first downstroke with a congruent upstroke. In plain figures, this means that the movement, instead of advancing from left to right, has been executed on the same spot, somewhat in the way in which soldiers pretend marching movements when "marking time," which involves the loss of at least $4/25$ of a second (18). The following *t* was written in the shape of an *l*, again; instead of advancing to the right, the writer went back to the left before writing the downstroke. The same phenomenon appears in the second downstroke of the *n*, which was written somewhat like an *e*. How strong the urge to the left (instead of to the right) was present in the writer is shown in the marked form of the *f* (four words further on) in the word *if*, where a pronounced movement to the left was added to the lower part of the character.

If, now, we follow downwards the very first characters of each line, we notice in any one case a strong retardation at the very start of the movement. The first word of the second line reads *others*. It stands in the middle of a sentence, and therefore the *o* is not meant as a capital letter. Again the writer wrote two strokes, of which the second covered the first, instead of writing an oval, and this means that the writing is very narrow, that there was a retardation because of the "marking time" movement and another retardation because the character was shaped in an angular form. We can, from these traces of writing-movement, establish that it took the writer $11/25$ of a second to shape her *o*, because we know that the covering stroke involves a loss of $4/25$ of a second, the angular form another of $2/25$, the returning to the left an additional $1/25$ of a second, which, added to the time of $4/25$ of a second needed for the normal shaping of the character, amounts to $11/25$ of a second (2).

I am not going into a more detailed measuring of time of the whole writing, because the reader will have no difficulty in spotting the numerous other instances of retardation and adjustments, and because we are not interested here in the actual figures, but rather in the fact that throughout the script again and again the writer could

not fully concentrate on her subject-matter, since she was almost permanently minding her calligraphy. To gather further evidence for our claim that this writing was deliberate and not the natural expression of the writer, we must catch her in numerous inconsistencies of her style. From what has been said already, we know that such inconsistencies show in the fact that a different style was applied at the beginning of words from that at their end. Again, a writer may emphasize the beginning of words to make them more striking, or under-emphasize them by making them less conspicuous. In the present case we can hardly doubt that we are confronted with what we call an *initial emphasis*. To make my concept perfectly clear, I have to beg the reader's pardon for describing this typical feature more at length than he might think necessary.

In the first word, *strong*, the *s* is evidently flourished and too large; we cannot expect the writer's endeavor completely to break down in so short a monosyllable, but still we notice that the *r* is much smaller than the *s*, and again the *o* smaller than the *r*; then the writer pulls herself together, and, in fact, manages to write a somewhat bigger *n*, immediately to fall back into her smaller style by writing a *g* with a loop which is smaller than the preceding *n*. The next word, *that*, starts with the *t* much bigger than the *t* at the end; the following word, *you*, ends with a *u* whose final downstroke has only $1/5$ of the length of the preceding *o*.

Our curiosity has by now been greatly aroused as to what the writer might do if she has to write a word of three syllables. We notice that such a task far surpasses her calligraphy reserves. She has to split the word *wondered* to take a rest in between and so to enjoy a break. The first part, *won*, shows a *w* whose initial downstroke is almost three times as long as the final downstroke of *n*. The rest of the word is written in the same inconsistent style. Again, the oval in the initial *d* is almost twice as big as in the final. Even in the short word, *the*, the last of the line, where the second part of the *h* is written as if an *e* were meant (because of the retarding movement to the left), this preceding "*e*" is bigger than the final *e*. In the first word of the second line, *others*, the *o* is far too big and conspicuous by its strange form, and, in fact, four times as big as the final *s*. The reader will now have no difficulty in tracing the same inconsistency in any one single word of the whole text.

In addition to the initial adjustments which prove the deliberation of the writing, and to the initial emphasis which proves the incon-

sistency of the style, the reader will easily establish the very slow tempo of the writing when checking its features with our table of indications of quick and slow writing movements. The reader will by now be also aware that we are not making our conclusions from some sporadic "graphological signs," but rather from whole groups of features which occur so regularly and frequently that they cannot possibly be ascribed to chance causes.

What, now, is the social behavior of this writer? She evidently plays a part, wants to impress others by some virtues which she pretends to have because of her own vain feeling that her plainness would be found out if she gave up her pose. She plays the part of a striking and original personality, but is evidently not up to this self-imposed task. Her doubts about her adequacy are so strong as to make her suddenly stop in the middle of her loquaciousness. In the middle of the third line she even goes so far as to retrace her letters, though her pen works quite properly, and though her retracing impaired rather than improved the legibility of the word *know*. Even when proceeding to the next word, *would*, she feels so uncertain of herself that her subconscious doubt stopped her movement for, say, 8/25 of a second to write a slanting covering stroke instead of an oval in the word *would*. In the same word she reduced the ratio in the size of her characters so as to write an *l* no bigger than a normal *e*. And in the next line just beneath in the word *will* the two *l*'s are written as if small and not medium letters were meant.

Any reduction of the size-ratio impairs the legibility of the script. In a quick writing such a reduction is only a natural consequence of the writer's urge to produce the greatest possible result with the smallest possible effort, but in a slow and deliberately shaped writing the same feature is due to the subconscious urge for non-committal expression.

To estimate the features of the handwriting in Figure 11 correctly, one has to know that the writer is not older than 30, and enjoys good health. Being given this information, the reader will wonder how it comes about that a young, healthy man of complete graphic maturity writes so slowly and hesitatingly, in fact, with rather timid movements. Examined with the magnifying glass, the strokes appear trembling and broken. The bizarre shape of the symbol for \mathfrak{G} , and the similar shaping of the final loops added to the *t*, show a marked retardation. Other strokes have been turned to the left, e.g., the very final stroke in the *y* in the last word.

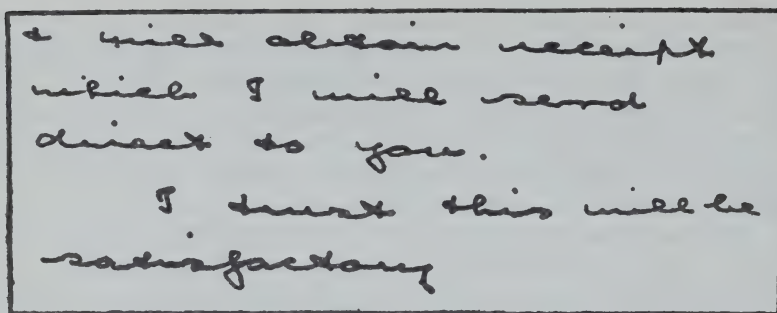


FIGURE 11

Furthermore, we notice that the grasp of the pen was very loose indeed, that there is hardly any difference in the shadowing of up- and down-strokes, or, anyhow, of the strokes running in different directions, which would necessarily be the case if a writer with a properly working musculature firmly grasped his pen.

In spite of the very slow, timid movement, the writer cannot perform a single syllable without taking a rest in between. In the first word of the third line, *direct*, there is a dot, in fact, a resting point, between the *i* and the *r* within the stroke, and the same happened, for instance, in the last word, *satisfactory*, between the *a* and the *t* and between the *i* and the *s*. Also in this writing the shapes of the characters are neglected or rather skimmed, and quite different characters are written in the same way, e.g., *w* like *m* in the word *will*; the *r*'s are always written as if *i*'s were meant.

In addition to all these features of unnaturalness, which prove that the writer feels intimidated and inferior towards the outer world, we notice strong initial *under-emphasis*, and the *I*, the symbol of the writer's Ego, is also under-emphasized and written with the same uncertainty to which the retraced and touched-up characters are due.

This writer in his social behavior tries to make himself inconspicuous, so as not to attract the attention of those around him. From the graphological criteria he is a hypocrite with a "bad conscience." He happens to be an employee whose non-integrity was realized by his employers when his actual frauds were found out after I had given an S. O. S. concerning his trustworthiness.

REFERENCES

1. BINET, A., & COURTIER, —. Sur la vitesse des mouvements graphiques. *Rev. phil.*, 1893, **35**.
2. BROOKS, C. H. Your character from your handwriting. New York: Morrow, 1930.
3. CRÉPIEUX-JAMIN, J. L'écriture et le caractère. Paris: Alcan, 1889.
4. ———. Les éléments de l'écriture des canailles. Paris: Flammarion.
5. ———. A B C de la graphologie. Paris: Alcan, 1929.
6. DREVER, J. The analytical study of the mechanism of writing. Edinburgh, 1914.
7. FREEMAN, F. N. The handwriting movement. Chicago,
8. JUDD, C. H. Genetic psychology for teachers. New York, 1907.
9. KAUFMAN, J. H. Schreibversuche mit Volksschülern.
10. KLAGES, L. Die Probleme der Graphologie. Leipzig: Barth, 1910. Pp. xii+260.
11. ———. Handschrift und Charakter. Leipzig, 1916.
12. McALLISTER, C. N. Research on movement used in writing. *Stud. Yale Psychol. Lab.*, 1900, **8**.
13. MEUMANN, E. Vorlesungen zur Einführung in die experimentelle Pädagogik. Leipzig: Englemann, 1915.
14. MEYER, G. Die wissenschaftlichen Grundlagen der Graphologie. Jena: Fischer, 1901.
15. MICHON, J. H. Système de Graphologie. Paris, 1875.
16. PREYER, W. Zur Psychologie des Schreibens. Hamburg: Voss, 1895. Pp. 230.
17. SAUDEK, R. Psychology of handwriting. New York: Doran, 1926. Pp. 288.
18. ———. Experiments with handwriting. New York: Morrow, 1929. Pp. 470.
19. ———. The methods of graphology. *Brit. J. Med. Psychol.*, 1927, **7**, 221-259.
20. ———. Experimental graphology. *Sci. Prog.*, 1929, **23**, 468-486.
21. ———. Das zentrale Nervensystem und der Schreibakt. *Jahrb. d. Charakterol.*, 1929, **7**.
22. ———. Pathologische Merkmale in der Handschrift. *Soz. Med.*, 1929, **12**,
23. ———. Betonung und Unterbetonung von Wortanfängen und Wortenden als charakterologische und psychopathologische Merkmale der Schrift. *Zsch. f. angew. Psychol.*, 1930, **37**.
24. ———. Zur psychodiagnostischen Ausdeutung des Schreibdrucks. *Zsch. f. angew. Psychol.*, 1931, **39**.

19 Corringham Road
Golders Green
London, NW 11
England

LES MOUVEMENTS DE L'ÉCRITURE COMME INDICES DU COMPORTEMENT SOCIAL DE L'ÉCRIVAIN

(Résumé)

Afin d'établir si l'écriture peut, en quelque manière, révéler des traits positifs de la conduite sociale de celui qui écrit, et, s'il en est ainsi, de quelles particularités ou parties de l'écriture nous pouvons arriver à de telles conclusions, il a fallu examiner :

1. Quels facteurs déterminés coöperent à la formation de notre écriture.

La réponse est que nous ne savons pas encore le nombre exact de ces facteurs, mais que nous avons réussi expérimentalement à en isoler 15, et que, d'après l'écriture, nous pouvons dire lequel de ces facteurs prédomine dans une écriture donnée.

2. Si quelqu'un peut déguiser sa main habituelle à un tel point qu'un expert ne pourrait retrouver, sous ce déguisement, ses habitudes normales d'écriture.

La réponse est que presque personne ne peut cacher tous les signes caractéristiques de l'écriture simultanément, et que, tout au mieux, on ne réussira qu'en écrivant à une vitesse moindre que la normale; que le commencement d'un texte long serait mieux déguisé que la fin; qu'on réussirait mieux dans les parties en évidence et point du tout dans celles qui ne le seraient pas; que l'inconsistance du style présentées dans les parties évidentes et plus lentes d'une part, et les parties non en vue et plus rapides d'autre part, montreront la façon dont on contrôle son expression graphique.

3. Jusqu'à quel point le tempo du mouvement de l'écriture actuellement employé change l'apparence de l'écriture, et comment déterminer d'après tout ne se compose que des traces graphiques de ce mouvement, la vitesse dans tout passage particulier du texte.

On a établi que quand on écrit d'une façon qui n'est pas naturelle, nous ne pouvons écrire rapidement, parceque notre attention est partagée entre notre sujet et notre calligraphie. Nous pouvons certainement porter un jugement sur les traits caractéristiques déterminés ou sur la conduite sociale de celui qui écrit d'après les variations du tempo de l'écriture dans différents passages d'un long document, et des nombreuses inconsistances dans des traits définitifs de l'écriture.

SAUDEK

SCHREIBBEWEGUNGEN ALS MERKMALE DES SOZIALEN BENEHMENS DES SCHREIBER'S

(Referat)

Um zu ermitteln, ob und aus welchen bestimmten Merkmalen der Schrift erkannt werden kann, wie sich der Schrifturheber zu seiner Umgebung einstellt, musste zunächst untersucht werden:

1. Welche bestimmte Faktoren bei der Gestaltung unserer Schrift zusammenwirken.

Die Antwort darauf lautet, dass wir noch nicht alle diese Faktoren kennen, dass es uns aber gelungen ist, 15 solcher Faktoren experimentell zu isolieren, und dass wir erkennen können, welcher von diesen Faktoren bei der Entstehung einer gegebenen Schrift einen entscheidenden Einfluss ausgeübt hatte.

2. Ob es möglich ist, dass wir unsere habituelle Schrift so sehr will-

kürlich verändern, dass ein Schriftsachverständiger nicht in der Lage wäre, die normalen Schreibgewohnheiten aus der verstellten Schrift zu erkennen.

Die Antwort lautet, dass kaum jemand imstande ist, sämtliche Merkmale seiner normalen Schrift gleichzeitig zu verändern und dass es ihm teilweise nur dann gelingt, wenn er sein übliches Schreibtempo wesentlich herabsetzt; dass der Anfang eines längeren Textes gründlicher verstellt sein wird als der Schluss; dass es ihm besser gelingen wird, die auffälligen Schriftmerkmale zu verändern, dass er weniger Erfolg haben wird bei der Verstellung der weniger auffälligen Teile, und schliesslich dass er die unauffälligen überhaupt nicht verstellen wird; dass aus der Inkonsequenz der Stile in den auffälligen und langsam geschriebenen Teilen einerseits und den unauffälligen und schneller geschriebenen. Partien der Schrift andererseits darauf geschlossen werden kann, dass der Schreiber seine Ausdrucksbewegung aus dem Wunsch nach Effekt willkürlich beeinflusst hat.

3. Inwiefern das Tempo der Schreibbewegung das Schriftbild verändert und wie aus der Handschrift, die doch lediglich aus fixierten Spuren dieser Bewegungen besteht, die Geschwindigkeit, mit der jeder einzelne Teil der Schrift erzeugt wird, festgestellt werden kann.

Es ist bewiesen, dass ein unnatürlich Schreibender nicht schnell schreiben kann, weil seine Aufmerksamkeit teils auf den gedanklichen Inhalt des zu Schreibenden, und teils auf die äussere Form gerichtet ist. Aus den Veränderungen der Schreibgeschwindigkeit in den verschiedenen Teilen eines längeren Schriftstückes und aus den zahlreichen Inkonsequenzen bestimmter Schriftmerkmale kann man auf das soziale Verhalten des Schreibenden eindeutige Schlussfolgerungen ziehen.

SAUDEK

TEMPERAMENT AND RELIGIOUS EXPERIENCE*

From the Department of Psychology, University of Minnesota

KEITH SWARD

Classic descriptions of the "religious temperament" are to be found in the writings of James, Leuba, DeSanctis, and others, which indicate that religious experience, at least in its more extreme forms, is correlated with traits of a primarily emotional nature. The vocational choice of the typical mystic is regarded in this light not as an affair of the intellect, but rather as a function of non-intellectual drives or incentives. In particular, it is often stated that the tender-minded, shut-in type of individual is especially susceptible to religious appeals and frequently turns to religious pursuits in preference to the more customary walks of life.

Earlier attempts to psychologize religion have been based largely upon biographies of the Christian saints and mystics. While these investigations have raised a number of general problems, they have not provided a sufficiently objective account of the traits involved in religious experience or stated precisely to what class of religiously motivated individuals their descriptions are supposed to apply.

Profiting from recent advances in the field of personality measurement, the present study attempts to relate a single variety of religious experience to the concepts of "introversion" and the "inferiority complex." The general procedure was as follows: Self-ratings were obtained from 80 Roman Catholic divinity students on standard scales for measuring introversion (7) and inferiority attitudes (9). Comparative norms were established by presenting the same scales to college students, to university faculty members, and to a group of business men.

In anticipation of the results, it may be stated that the data afford striking confirmation of the testimony collected from saints and mystics of the Christian Church. The present findings, insofar as self-ratings constitute valid evidence, indicate that men selected into a Roman Catholic seminary are characterized to a marked degree by

*Recommended for publication by Walter R. Miles, accepted for publication by Carl Murchison of the Editorial Board, and received in the Editorial Office, August 20, 1930.

the emotional attitudes which are diagnostic of introversion and inferiority attitudes.

PROCEDURE AND SELECTION OF CASES

Self-ratings on the two scales in question were obtained from 80 divinity students enrolled in a Roman Catholic seminary in the Middle West. Over half of the group were familiar with testing procedure, and the rating was conducted as part of the course in psychology which they were taking at the time. The blanks were filled out anonymously during regular class periods under the supervision of the professor of philosophy at the seminary. No special instructions were given to the group except for the usual emphasis upon honesty and frankness in the ratings and a general statement that the data were to be used in a scientific study.

The university faculty members comprised the teaching staff of science departments at the University of Minnesota. Cooperation was sought by means of personal interviews with each member of the faculty who held the rank of an assistant professor or above. Sixty-eight out of 80 available subjects responded to the extent of sending in their ratings.

The 52 business and professional men included in the study were members of the St. Paul Kiwanis Club. The blanks were distributed to these subjects at the close of a regular weekly meeting. Approximately one-third of the club members cooperated.

Fairly standardized instructions were given to both of the latter groups concerning the supposed object of the investigation, which was described as a study of the traits of men in different occupations. In both cases the ratings were made anonymously and returned to the Psychology Department in self-addressed envelopes.

Previously published norms for college students were based on the ratings of 347 Dartmouth sophomores and the ratings of about 2000 freshmen who filled out the blanks as part of the Minnesota College Entrance Examination in 1926 and 1927 (4). Additional data were available for a group of 125 college upperclassmen who had been approached in connection with a study of extra-curricular activities at the University of Minnesota (12). Half of the students in the latter group were campus leaders; the remaining half had not participated in outside activities during their college careers. As neither total scores nor individual items differentiated campus leaders from control subjects, considered as total groups, these data were combined in the present study.

DESCRIPTION OF THE SCALES

The Heiddreder scales of introversion-extraversion and the inferiority complex were used in the study. The measure of introversion contains a list of 33 traits which were given by Freyd (6) as the distinguishing features of the condition (7). The scale intended as a measure of inferiority attitudes lists 93 traits which were culled from the psychiatric literature as outstanding symptoms of the so-called inferiority complex (9). In the form used for this study, both groups of traits were included on the same list, which was labelled "A Personal Traits Rating Scale."

Ratings were made on a five-point scale, marked off with the following steps: — — — 0 + ++. The double plus sign signified possession of a trait to a marked degree; the single plus column, its presence to a slight degree. The opposite of a trait to a slight or marked degree was indicated by checking the single minus or the double minus signs. Checking the 0 column was to indicate that neither the trait nor its opposite characterized the individual.

The measure of inferiority attitudes gives a coefficient of reliability of $+.73 \pm .03$ between original tests and retests (same form) of 147 college students after an interval of six weeks (9). Odd-even reliabilities were calculated from the present data. When corrected by the Spearman-Brown formula the values for the different sets of data range from $+.78 \pm .02$ to $+.90 \pm .02$ on the test of introversion, and from $+.84 \pm .02$ to $+.95 \pm .02$ on the inferiority attitudes scale.

Crucial tests of validity have not been applied to either measure by checking test scores directly against outside criteria. Some evidence exists, however, which throws indirect light on the question of validity. Low positive correlations were found between inferiority attitudes and two conditions which might conceivably be related to the trait, namely, physical defects and low economic status (4). The introversion scale, when applied to patients diagnosed as insane, confirms clinical expectations in distinguishing between manic-depressive and dementia praecox patients (1). It hardly needs to be pointed out, however, that whenever the terms "introversion" and the "inferiority complex" are applied in the discussion which follows, these labels are used provisionally.

TREATMENT OF THE DATA

According to the scoring scheme devised for each scale (7, 9), total scores consist of an algebraic sum of the plus and minus values

assigned to separate traits. Positive weights are given to items indicating possession of introversion or inferiority attitudes and negative values are assigned to items which indicate an absence or lack of the trait considered. Separate scores were computed for each scale and the results will be treated separately for each set of data.

The ratings were analyzed, first, by comparing measures of central tendency, and then by means of an item-by-item analysis of the blanks of seminary students and the ratings of 125 University of Minnesota undergraduates. The significance of the differences between mean scores of theology students and mean scores of comparison groups were determined in each case by computing the ratio of the difference to the probable error of the difference.¹ The item-by-item analysis utilizes exactly the same procedure except that here single items on both lists were examined one by one for their diagnostic value. For every trait the percentage of seminary students checking each symbol was compared with the percentage of college upperclassmen checking the same symbol. The same basic formula for evaluating group differences was then applied.

In presenting the data, results based on the comparison of mean scores will be given first, followed by the detailed analysis of individual traits which proved diagnostic on the two lists.

Means, standard deviations, and the differences in central tendency based on the introversion scale are shown in Table 1.

RESULTS

Results Based on the Introversion Scale. Perhaps the first point to be noted is that the mean score of every group except divinity students has a low negative or minus value,² indicating that the normal or average condition, at least for the groups represented, is a slight tendency towards the absence of introversion. In contrast to the results for college students, business men, and faculty members, the average score of theology students is positive, lying 13 to 15 points higher on the scale in the direction of introversion. Possession of the

¹The formula used for computing the probable error of the difference was as follows:

$$P.E._{diff.} = \sqrt{P.E._{av. 1}^{(2)} + P.E._{av. 2}^{(2)}}$$

²Positive scores signify possession of traits associated with introversion; negative scores, an absence or lack of the condition.

TABLE 1

MEANS, STANDARD DEVIATIONS, AND PROBABLE ERRORS OF THE MEANS FOR RATINGS ON INTROVERSION; ALSO THE DIFFERENCES IN CENTRAL TENDENCY BETWEEN SEMINARY STUDENTS AND NON-DIVINITY GROUPS

The ratings	No. of cases	<i>Av.</i>	<i>S.D.</i>	<i>P.E.</i> <i>av.</i>
Theology students	80	7.2	14.5	1.1
College upperclassmen	125	—6.6	13.3	0.8
Freshman men (Twin City)	250	—8.2	13.8	0.6
Freshman men (non-Twin City)	250	—4.4	16.7	0.7
Freshman women (Twin City)	250	—8.1	16.9	0.7
Freshman women (non-Twin City)	233	—5.8	15.3	0.7
Business men	52	—6.6	16.2	1.5
Faculty members	68	—6.9	14.3	1.2

The differences

	<i>Diff.</i>	<i>P.E.</i> <i>diff.</i>	<i>Diff.</i> <i>P.E. diff.</i>
College upperclassmen	13.8	1.36	10.2
Freshman men (Twin City)	15.4	1.25	12.3
Freshman men (non-Twin City)	11.6	1.30	8.9
Freshman women (Twin City)	15.3	1.30	11.8
Freshman women (non-Twin City)	13.0	1.30	10.0
Business men	13.8	1.86	7.4
Faculty members	14.1	1.63	8.7

trait seems to be the characteristic condition of seminary students as a group.

Apparently the emotional attitudes measured by such ratings are decidedly more prevalent among divinity students than among unselected individuals, irrespective of the group which one considers as a norm. All of the resulting differences in central tendency shown in Table 1 are statistically significant.³ The lowest critical ratio is 7.4.

Overlapping shows the same differentiations. The percentage of ratings of divinity students reaching or exceeding the mean of each comparison group is as follows: 80.8, 82.8, 77.1, 82.6, 79.8, 80.8, and 81.1. These values are presented in the same order as the differences in Table 1.

³Garrett states that an obtained difference between means must be at least four times the probable error of the difference to be considered statistically significant (5, p. 136).

In spite of highly significant differences in central tendency, however, the amount of variability in the distribution of the trait, expressed in terms of standard deviations, is practically the same for all groups.

Results Based on the Inferiority Complex Scale. The results on inferiority attitudes given in Table 2 corroborate the findings on introversion. Emotional attitudes in general, regardless of the particular pattern to which they belong, are more characteristic of theology students. On this scale the mean score of every group, but one, deviates in a positive direction,⁴ indicating that the average condition here, as contrasted with results on introversion, tends to be possession of traits related to the inferiority complex. One group stands out as an exception to the general rule. The ratings of faculty members give a mean of minus 12.8.⁵

A further contrast between these results and those based on introversion arises from the fact that theology students do not constitute a homogeneous group with respect to inferiority attitudes. There is a significant difference in mean score between the groups designated in the table as seminary juniors and advanced theology students.⁶ Because of this discrepancy the results will be treated for each group separately. It is interesting to note, in passing, that on the scale for introversion the means for the two groups of divinity students were indistinguishable. Possible reasons for the disagreement of the tests at this point will be cited after presenting the data as it stands.

Three sets of data will be given (see Table 2), corresponding to the comparisons of non-theology groups with (1) seminary juniors, (2) advanced theology students, and (3) the total group of divinity students. The results will be taken up in this order.

⁴Positive scores indicate possession of inferiority traits; and negative scores, a tendency towards the opposite of the condition.

⁵It might be argued that the results obtained from college professors are not "genuine" in that the group assumed a sophisticated attitude toward the scale and refused to commit themselves. While it is impossible to rule out this supposition altogether, it should be noted that the conditions of rating were expressly designed to counteract such an attitude; and if this factor operates at all, it ought to influence ratings on the other scale as well. But the traits measuring introversion, though listed on the same rating scale, do not differentiate faculty members from other non-seminary groups.

⁶The seminary is organized on the following plan: Its six-year curriculum consists of a four-year academic college course, followed by two years of theological training. Students in both divisions are candidates for the priesthood. The seminary juniors were third-year students enrolled in the lower division. The advanced theology students were completing their final year of study in the upper division of the institution.

TABLE 2

AVERAGES, STANDARD DEVIATIONS, AND PROBABLE ERRORS OF AVERAGES FOR RATINGS ON INFERIORITY ATTITUDES; ALSO THE DIFFERENCES IN CENTRAL TENDENCY BETWEEN THEOLOGY STUDENTS AND NON-SEMINARY GROUPS

The Ratings	No. of cases	Av.	S.D.	P.E. av.
Seminary juniors	44	52.0	37.4	3.8
Advanced theology students	36	30.8	34.3	3.9
Total divinity group	80	43.8	38.6	2.9
Freshman men	1558	32.9	43.1	0.7
Freshman women	1258	38.2	40.8	0.8
Dartmouth sophomores	347	17.9	42.8	1.5
College upperclassmen	125	4.9	43.2	2.6
Business men	52	6.2	45.6	4.3
Faculty members	68	-12.8	39.4	3.2

*The differences
For seminary juniors*

	Diff.	P.E. diff.	Diff. P.E. diff.
Advanced theology students	21.2	5.45	3.9
Freshman men	19.1	3.86	5.0
Freshman women	13.8	3.88	3.6
Dartmouth sophomores	34.1	4.09	8.3
College upperclassmen	47.1	4.60	10.2
Business men	45.8	5.74	8.0
Faculty members	64.8	4.97	13.0

For advanced theology students

Freshman men	2.1*	3.96	0.5
Freshman women	7.4*	3.98	1.9
Dartmouth sophomores	12.9	4.18	3.1
College upperclassmen	25.9	4.69	5.5
Business men	24.6	5.81	4.2
Faculty members	43.6	3.93	11.1

For the total divinity group

Freshman men	10.9	2.98	3.7
Freshman women	5.6	3.01	1.9
Dartmouth sophomores	25.9	3.26	7.9
College upperclassmen	38.9	3.89	10.0
Business men	37.6	5.19	7.2
Faculty members	56.6	4.32	13.1

*Differences contrary to the characteristic direction.

1. Consistently large and significant differences are found between the mean score of seminary juniors and the average scores of non-divinity groups, indicating relatively higher incidence of inferiority attitudes within the theology group. These comparisons comprise the first set of differences in Table 2. Only one of the critical ratios is lower than 4.0, i.e., in the case of the comparison with freshman women at the University of Minnesota the reliability ratio is 3.6.

The amount of overlapping reveals the same trend in the data. The percentage of the ratings of seminary juniors reaching or exceeding the averages of other groups in the direction of inferiority, when arranged in the same order as the first list of differences in Table 2, is as follows: 69.1, 66.7, 61.4, 76.0, 82.9, 82.1, and 96.7.

Self-ratings of 44 individuals who are third-year candidates for the priesthood in a Roman Catholic seminary, then, show generally significant differences in central tendency when compared with the ratings of college students, faculty members, and business men. Theology students record much more frequently the possession of traits which are symptomatic of the inferiority complex.

2. Inferiority attitudes are not as conspicuous among advanced theology students as among younger individuals in the same institution. Hence, the second set of results in Table 2 are neither as striking nor as consistent as is the case with seminary juniors. Two of the differences—for comparisons with Minnesota freshmen of both sexes—although not statistically reliable, actually run counter to the characteristic direction. The differentiation between Dartmouth sophomores and the more mature seminary group, although confirming the general trend of the data, falls just short of being statistically significant.

Similar comparisons may be stated in terms of overlapping. These indices, expressed as before, and arranged in the same order as the second list of differences in Table 2, are as follows: 46.0, 38.7, 61.7, 75.0, 72.3, and 92.4.

With the exception of comparisons with college freshmen, the ratings of 36 advanced theology students confirm the results obtained from seminary juniors, namely, that traits associated with the inferiority complex are present to a more noticeable degree among men preparing for the priesthood than among people generally.

3. The third and last set of results in Table 2 was obtained by combining the scores for both groups of divinity students. All but

two of the resulting differences are statistically significant, and all are in the direction indicating greater prevalence of the trait among divinity students as a total group.

The indices of overlapping which correspond to the last set of comparisons in the table are as follows: 58.7, 54.1, 69.5, 79.5, 77.7, and 94.6.

Summarizing the results on this scale, one may say that: (*a*) inferiority feelings are definitely more common among young theology students than among individuals selected from the general population; (*b*) relatively more mature seminary students, on the other hand, are not subject to the condition to so marked a degree. In fact, college freshmen in a state university are characterized by the trait to a greater extent than are advanced theology students, although the differences which oppose the general direction of the findings are not statistically significant.

DISCUSSION OF DIFFERENCES IN CENTRAL TENDENCY

Two inconsistencies in the data should be mentioned at this point: first, the discrepancy between the two groups of theology students on inferiority attitudes; and, second, the disagreement at several points between results derived from the two scales.

In connection with these seeming irregularities in the results, it should be observed that ratings on inferiority are apparently related to age, while the same fact does not hold for ratings on introversion. Several lines of evidence suggest a relationship between inferiority feelings and maturity: (*a*) The scale distinguishes between freshmen at the University of Minnesota and upperclassmen in the same institution. Freshmen have consistently higher scores in the direction of inferiority, with critical ratios of the differences between means of 12.2 and 10.4. (*b*) The lack of homogeneity within the group of divinity students has been pointed out. This discrepancy may be partly a function of maturity. (*c*) Faculty members have decidedly lower scores, in the direction of non-inferiority, than do less mature groups of college age. The differences here are uniformly large, yielding reliability ratios from 4.3 to 15.5. (*d*) Similarly, business men are less subject to the condition than are younger individuals. The differences in this case give critical ratios ranging from 3.6 to 8.0. Only one difference, derived from the comparison of business men with college upperclassmen, runs counter to the direction supposedly expected on the basis of age.

The significance of the age factor is another question. On the one hand, inferiority attitudes may actually decrease with increasing age. Or it may mean that mature individuals when confronted with a rating scale are simply less naïve and less willing to commit themselves on the possession of traits which are commonly recognized as socially undesirable. But the supposed age difference *does not occur* for the traits measuring introversion, and these were included on the same list and rated at the same time.

If one may assume that the relation between the inferiority complex and maturity is a genuine phenomenon and has its counterpart in changes that occur in actual behavior, the inconsistencies noted in the data may be largely accounted for on this basis.

1. In the first place, the disparity between the ratings of advanced theology students and less mature divinity students may be a function of age. The difference in mean age of three or four years⁷ may have a bearing on the dissimilarity between the two groups in temperament. Conceivably, a critical period of adjustment could be partially responsible for the higher incidence of inferiority traits among less mature candidates for the priesthood.

2. Closely related to the fact that differences in age may account for the discrepancy between the two groups of seminary students on inferiority attitudes is the inconsistency of results given by the two scales. Groups differing widely in age are not discriminated by the measure of introversion on the basis of maturity, and on this scale theology students were more or less homogeneous in their ratings. Here, age does not seem to be a factor. With the inferiority scale a different situation occurs. Whereas comparisons with seminary juniors are all in the expected direction and give fairly large differences, advanced theology students are less sharply marked off from groups representing the general population; and younger subjects on the whole, i.e., college freshmen, show the closest resemblance to the advanced theology group. If age is a factor affecting the prevalence of the trait, one would expect its more frequent occurrence among college freshmen than among college juniors and seniors.

7

	Mean age in years	S.D.
Seminary juniors	23.4	1.9
Advanced theology students	26.8	3.4

The latter, who are more comparable in age to the group of mature divinity students, and belong to the same sex, show a significantly lower incidence of the trait than do individuals undergoing religious training.

Sex differences (4, 9) are probably at work, also, in minimizing the differences between women college students and theology students. It will be recalled that freshman women possess the trait to even a greater degree than do advanced theology students, although the amount of overlapping is large. In the case of freshman women, then, it is likely that two factors operate to increase the scores of the group in the direction of inferiority, i.e., both the effect of age and the fact of sex differences in the trait. However, when freshman women are compared with younger seminary students, although the latter are approximately five or six years older on the average, the greater prevalence of the trait among divinity students is quite apparent. That is, the supposed correlation between temperament and religious experience emerges in spite of the sex difference and the presumed influence of maturity.

Therefore, the explanation of two irregularities in the data, i.e., the lack of homogeneity among theology students as regards inferiority attitudes and the disagreement of the two scales, goes back in all probability to the fact that *other factors* related to the trait break through, obscuring somewhat the relationship between religious experience and temperament.

Intercorrelations of the Ratings. To determine to what extent introversion and inferiority attitudes were independent of one another, the two measures were correlated for each set of data, using in each case the product-moment formula. The intercorrelations are given in Table 3.

The substantial degree of correlation between the two measures would imply that, whatever traits are measured by the respective scales, they are not discrete, independent features of the personality, but represent attitudes that fuse and overlap more or less and, taken together, indicate a general instability or emotionality rather than specific patterns. Further evidence along the same line was obtained by giving an additional test of emotionality, the Woodworth Psychoneurotic Inventory, to the group of 125 college upperclassmen. For these subjects, the correlation between ratings on introversion and scores on the Woodworth Questionnaire was $+.51 \pm .04$; and between the Woodworth Inventory and the inferiority scale, $+.59 \pm .04$.

TABLE 3
CORRELATIONS BETWEEN RATINGS ON INTROVERSION AND RATINGS ON THE
INFERIORITY COMPLEX

	No. of cases	<i>r</i>	<i>P.E.</i> <i>r</i>
Seminary students	80	+.57	±.05
Faculty members	68	+.38	±.07
Business men	52	+.45	±.08
College upperclassmen	125	+.59	±.04

For the purpose of this study it is probably sufficient to state that a general "tender-mindedness" characterizes the divinity group, according to their subjective report. At any rate, whether the results are to be regarded as evidence of differences in particular traits or as evidence of dissimilarity in a more general factor of emotionality, the two scales agree in the general trend of the differences which occur. Both sets of ratings on temperament discriminate between people in general and individuals preparing for the priesthood.

The item analysis of the ratings, presented next, throws additional light on the particular pattern of traits characterizing seminary students.

Item Analysis of the Ratings. Perhaps the most important method of treating the data is the detailed comparison of the ratings of divinity students and the ratings of college upperclassmen, item for item, because by so doing one may answer a serious objection which might be lodged against the findings, namely, that other factors, apart from genuine differences in temperament, might be invoked as a plausible explanation of such differences as occur in the ratings.

Two extraneous influences which may have forced the results, irrespective of personality differences as such, deserve mention at this point. To what extent these factors constitute an artificial loading of the results one may judge from the following discussion and from the actual list of items given below.

1. In the first place, the ratings of seminary students may reflect a peculiar form of motivation. This group may have construed the scale as a test of "character," with a strained effort to appropriate traits in line with traditional religious standards. The divinity students may have simply committed themselves more freely or adopted in their ratings a semi-"confessional" attitude. Such a mental set, temporarily induced, could produce the apparent effect of significant

differences to a spuriously high degree and could have arisen as a function of the rating situation itself or as an habitual attitude or frame of mind characterizing the group.

It is conceivable that the ordinary social pressures and conventions to which college students, faculty members, and business men are subjected act, not only to repress *overt* expression of traits indicative of introversion and inferiority attitudes, but even to the point of inhibiting mere verbal admission as regards their possession. More "normal" groups may differ from individuals enrolled in a seminary, then, not only in the degree to which they display these traits in everyday behavior, but in the degree to which they are willing to *admit* that they have such traits. To what extent religious practices in the Catholic Church and the general régime of seminary life give rise to an habitual attitude of the sort described, it is impossible to say.

But, insofar as it was possible to control the immediate circumstances of rating, pains were taken to minimize the effect of an unusual or artificial attitude within the group. As described above, standard instructions preceded the anonymous rating process. The usual request was made that the ratings be frank and accurate. It was pointed out that the data were desired for a scientific study. Nothing was stated that might have made the group unduly self-conscious. Furthermore, the seminary juniors were familiar with testing procedure, having been subjected to mental tests earlier in the semester. It is interesting to note that this very group gives the most striking results on one of the scales. As far as the specific conditions are concerned, under which the rating process took place, it would seem that nothing intrinsic to the situation operated to induce an unnaturally introspective or forced attitude on the part of the theology group.

2. It might be urged, further, that the tendency among young priests towards introversion and feelings of inadequacy as expressed on a rating scale, is to some extent the result of conformity to conventions or standards which happen to have the stamp of approval within such a social group. In other words, the diagnostic traits might fit in with a stereotype that influences not only the behavior of divinity students, but their conscious preferences as well. Excessive humility, perfectionist tendencies, feelings of insignificance, and the like, might be primarily an expression of what James called "fashions of feeling" prevalent among certain Christian sects. If the ratings

are to be so explained and the distinguishing traits represent systematized attitudes which are deliberately cultivated and fostered by the Church, one is faced with the possibility that the results, whether due to actual differences in behavior or merely to differences in conscious or unconscious preferences, are primarily a function of the social pressures brought to bear upon men selected into an institution of the sort in question. How extensively these personality traits conform to traditional stereotypes "in vogue" among Catholic and Christian bodies generally, one may judge from an examination of the list itself. Tables 4 and 5 give the separate items on both scales which discriminate between seminary students and University of Minnesota upperclassmen.⁸ The method used for the item-by-item analysis was described in an earlier section.

As might be expected, the specific traits which characterize theology students merely corroborate in greater detail the results based on the comparison of central tendencies in the data. Table 4, giving the results on introversion, lists only "positively" diagnostic items, i.e., those items which characterize divinity students. "Negatively" diagnostic traits, or those characterizing college students, were in all cases merely the opposites of the traits listed above, or, occasionally, the position on the scale which signified neither presence nor absence of the traits which were diagnostic for seminary students. An item was included on the list if the difference between the percentage of theology students and the percentage of college undergraduates checking the same symbol yielded a critical ratio of 3 or greater. The first eight items in the table, arranged in order of decreasing diagnostic value, give critical ratios ranging from 7.2 to 5.0; items 9 to 14, ratios between 4.9 and 4.0; while the last nine items have diagnostic values between 3.8 and 3.1. Twenty-three or approximately two-thirds of the total number of items on the scale are diagnostic according to the test used.

Several items on the list, however, are not to be regarded primarily as evidence of introverted tendencies among theology students. At least, other explanations of their occurrence in the group are equally plausible, as opposed to the explanation on the basis of temperament. Two traits might be considered a function of the general restrictions

⁸College upperclassmen were chosen for the comparison as most comparable to the seminary group in general make-up. Both groups are of the same sex and of approximately the same age, although the divinity students probably run two to three years older on the average.

of seminary life, altogether apart from temperamental inclinations. These refer to the fact of "limiting acquaintances to members of one's own sex" (No. 6) and of "limiting one's acquaintances to a select few" (No. 20). Conformity to a stereotype or to prevalent

TABLE 4

ITEM-BY-ITEM ANALYSIS OF RATINGS ON INTROVERSION: INDIVIDUAL TRAITS WHICH CHARACTERIZED DIVINITY STUDENTS, ARRANGED IN ORDER OF DECREASING DIAGNOSTIC VALUE

-
1. Shrinks when facing a crisis. +*
 2. Is a poor loser. +
 3. Prefers to work alone rather than with people; prefers to work at tasks that do not bring him into contact with people. ++
 4. Is suspicious of the motives of others. +
 5. Shrinks from actions which demand initiative and nerve. +
 6. Limits his acquaintances to members of his own sex. +
 7. Is outspoken; says what he considers the truth regardless of how others may take it. —
 8. Hesitates in making decisions on ordinary questions in the course of the day. +
 9. Is reticent and retiring; does not talk spontaneously. +
 10. Avoids all occasions of talking before crowds; finds it difficult to express himself. ++
 11. Gets rattled easily; loses his head in excitement or moments of stress. ++
 12. Keeps in the background on social occasions; avoids leadership at social affairs and entertainments. ++
 13. Prefers to read a thing rather than to experience it. ++
 14. Dislikes and avoids any process of selling or persuading anyone to adopt a certain point of view (except in the religious field). +
 15. Works by fits and starts. ++
 16. Is extremely careful about the friends he makes; must know a person pretty thoroughly before calling him a friend. +
 17. Expresses himself better in writing than in speech. ++
 18. Prefers participation in competitive intellectual amusements to athletic games. +
 19. Introspects; turns his attention inward toward himself. ++
 20. Limits his acquaintances to a select few (this may be beyond his control). +
 21. Rewrites his social letters before mailing them. ++
 22. Keeps a diary. ++
 23. Resists discipline and orders. —
-

*The symbols placed after each trait indicate which particular step on the five-point scale, or what degree of a given trait, proved diagnostic in the ratings.

Christian standards could probably account for the presence of traits which indicate that seminary students are not "outspoken; speaking the truth regardless of how others may take it" (No. 7) and do not "resist discipline and orders" (No. 15).

The point in noting the consistency of the list is that, on the whole, these items of behavior cannot be harmonized with standards of conduct sanctioned by the Church or attributed to the influence of religious conventions rather than to the fact of personality differences as such. While four of the separate items can be partially or wholly explained away on this basis and a number of others may be susceptible to the same interpretation, the great majority strongly indicate that the temperament of divinity students exhibits in detail the pattern of traits typical of the so-called introvert. And this resemblance does not seem to be a mere product of atypical standards within the religious group. A similar consistency may be noted for

TABLE 5

ITEM-BY-ITEM ANALYSIS OF RATINGS ON INFERIORITY ATTITUDES; SEPARATE TRAITS WHICH CHARACTERIZE THEOLOGY STUDENTS, ARRANGED IN ORDER OF DECREASING SIGNIFICANCE OF THE DIFFERENCE

-
- | | | |
|-----|--|----|
| 1. | Has given serious thought to the question of his life work. | ++ |
| 4. | Feels nervous and inadequate when he knows that the thing he is doing is important. | ++ |
| 2. | Measures his abilities and achievements according to an ideal. | + |
| 3. | Is uncomfortable and ill at ease with people who are socially prominent or otherwise distinguished. | ++ |
| 5. | Wishes to do some one thing supremely well. | ++ |
| 6. | Is content with his lot. | ++ |
| 7. | Is given to remorse and regrets. | + |
| 8. | Applies advice, sermons, inspirational literature, etc., to himself. | ++ |
| 9. | Is often depressed. | + |
| 10. | Is sometimes oppressed in the midst of an enterprise by a sense of his unfitness to carry it through. | + |
| 11. | Uses mottoes, slogans, passages from favorite authors to keep up his spirits. | ++ |
| 12. | Is moved to sympathy by the misfortunes of persons he does not like. | + |
| 13. | Is moved to sympathy by the misfortunes of his friends. | ++ |
| 14. | Has high standards of work; wants his work to be distinctly above the average. | + |
| 15. | Uses good practical judgment in conducting his everyday affairs. | + |
| 16. | Feels depressed or resentful because of the advancement—social, professional, or otherwise—of persons who used to be his equals. | + |
-

TABLE 5 (*continued*)

ITEM-BY-ITEM ANALYSIS OF RATINGS ON INFERIORITY ATTITUDES; SEPARATE TRAITS WHICH CHARACTERIZE THEOLOGY STUDENTS, ARRANGED IN ORDER OF DECREASING SIGNIFICANCE OF THE DIFFERENCE

-
17. His feelings are easily hurt; is touchy, oversensitive. +
 18. Is self-conscious. ++
 19. Has experienced a humiliation which has affected him seriously. ++
 20. Is likely to be unappreciated. —
 21. Is sometimes overcome in the midst of an activity by a sense of its uselessness. ++
 22. Hesitates to put his abilities to a test. +
 23. Is dissatisfied with his progress and achievements up to the present time. —
 24. Is disturbed if compelled to spend time in disorderly, ill-kept surroundings. ++
 25. Is easily embarrassed. ++
 26. Often feels that something he has said or done has hurt someone's feelings or made enemies. ++
 27. Goes straight after what he wants; doesn't care who knows it. —
 28. Finds it hard to get over punishment; feels resentful and humiliated for a long time. ++
 29. Is sensitive to blame. ++
 30. Generally accomplishes a good day's work in a day. —
 31. Is embarrassed by the memory of "scenes" and blunders long after they have happened. ++
 32. Indulges in self-pity. +
 33. Is uncomfortable with people who are queerly dressed or whose manners are not correct. ++
 34. Doubts the wisdom of his decisions after they are made. ++
 35. Is misunderstood by most people. —
 36. Blushes readily. ++
 37. Hesitates to start things which he considers *important*; keeps putting them off. ++
 38. Enjoys taking part in an argument. —
 39. Feels detached and alone. —
 40. Has a pleasing personality. —
 41. Thinks that people usually get the rewards they deserve in this world. —
 42. Has social poise; is at ease in most social situations. —
 43. Is conscientious about his work; will do too much rather than too little. +
 44. Is considerably upset by defeat. ++
 45. Worries about his ability to succeed in fields where he most wishes to succeed. ++
 46. Wishes to live his life fully and completely—to express all his possibilities. +
-

the results in Table 5 based on the inferiority scale, although conventional attitudes seem to play a more conspicuous rôle here than on ratings for introversion.

The general arrangement of Table 5 and the criterion for inclusion of items are the same as for Table 4. The critical ratios range from 7.7 to 5.1 for the first nine items; from 4.9 to 4.1 for items 10 to 19; and from 3.9 to 3.0 for the last 26 items on the list.

Almost one-half of the traits in this table, however, are somewhat ambiguous in the sense that it is largely a question of interpretation whether their occurrence should be attributed to temperament or to the effect of social pressures. Of course, these interpretations are not mutually exclusive.

Quite a number of the traits are closely related to "perfectionist" tendencies,⁹ bringing out the point that theology students have "high standards of work," "wish to do some one thing supremely well," "are conscientious," "worry about their abilities to succeed," and the like. This cluster of items could be attributed to the general ideals or conditions to which the group is subjected rather than to any fundamental personality difference which gave rise to religious experience in the first place.

A number of other qualities might be equally well explained as rather typical standardized attitudes to be expected in such a group and probably afford less evidence of genuine temperamental differences. These refer to the fact that the group "applies advice, sermons, etc., to themselves" (No. 8), "uses mottoes, slogans, etc., to keep up their spirits" (No. 11), "is moved to sympathy by the misfortunes of others" (Nos. 12 and 13), and "is afraid of hurting other people's feelings" (No. 26). One attitude very definitely reflects a specific belief of the group, namely, that people fail "to get the rewards they deserve in this world" (No. 41).

That divinity students are "content with their lot" (No. 6), do not feel "detached and alone" (No. 39) nor "unappreciated" (No. 20), and are not "dissatisfied with progress and achievements up to the present time" (No. 23) seems to contradict the general trend of the data. On the other hand, one might be safe in assuming that these traits not only fit in with a traditional attitude but may have compensatory significance, judging from the tenor of the remaining items which imply that the group as a whole, according to their own

⁹Particularly the following items: 1, 2, 5, 14, 43, 45, and 46.

testimony, is anything but successful or well adjusted in a social sense.

Two traits are out-and-out exceptions to the general drift of the items. These have to do with the fact that the group "uses good practical judgment in conducting everyday affairs" (No. 15) and is not "misunderstood by most people" (No. 35). Possession of these traits to a *slight* degree characterizes theology students. On both traits, however, the direction of the difference is reversed as regards their presence to a *marked* degree. That is, while individuals in the seminary possess these traits to a slight degree, the double plus symbol on the same traits is "negatively" diagnostic or characteristic of university students.

Exclusive of the traits which are equivocal in that they are symptomatic of the inferiority complex and at the same time conform to patterns of behavior fostered by the Catholic religion, fully one-half of the list are sufficiently clear-cut to establish a relationship between religious experience and inferiority attitudes as such. Taken at their face value, both sets of ratings, when subjected to an item-by-item analysis, produce what appear to be true differences in temperament.

MOTIVATION OF RELIGIOUS EXPERIENCE

All that one is in a position to say concerning the findings presented thus far is that feelings of inferiority and introverted tendencies, as indicated by self-ratings, are considerably more prevalent among theology students than among other individuals.

This conclusion rests, of course, upon two assumptions: first, that ratings are valid indications of traits as they are manifested in actual behavior; and, second, that the norms used for this study represent a fair cross-section of society. It may be that norms based on a random sample of persons of the same age drawn from the population at large might not substantiate the results obtained in the present study.

If these suppositions are at all tenable, it appears that emotionality in the form of introversion and inferiority attitudes may be regarded as a condition which predisposes individuals toward religious experience. The vocational choice of candidates for the Roman Catholic priesthood, therefore, may be in part a function of emotional drives which divert such individuals from the more usual social channels. This temperamental susceptibility is probably best described as a general "tender-mindedness" rather than as highly specific patterns

of traits separable by hard and fast distinctions. That such a scheme of motivation applies to the present group is borne out by the evidence which James and Leuba have culled from the lives and confessions of great religious characters.

Certain reservations should be mentioned, however, in connection with the interpretation of religious experience offered above. (a) Irrespective of the adequacy of the data to explain personality factors involved in the adjustment of the present group of theological students, no inferences may be drawn from the data regarding the motivation of religious experience in the lives of normally adjusted individuals or within widely differing religious sects. (b) Secondly, these conclusions, if warranted, are not offered as an all-sufficient account of religious motivation even within the group concerned. Undoubtedly, such behavior is conditioned by a multiplicity of factors and cannot be reduced to a single cause or ascribed to the operation of one mechanism exclusively. Temperamental predispositions constitute but one more or less conspicuous element. (c) Furthermore, no claim is made for the specificity of religious experience in relation to whatever temperamental inclinations are indicated by the data. If selection of the priesthood is to some extent a function of "tender-mindedness," it represents but one of the possible outlets or channels which the temperament in question might have sought out. It may be largely a matter of accident that the present group, if so constituted, turned to religious experience rather than to other characteristic forms of expression. This choice may have been in some measure a fortuitous consequence of the fact that the early environment of the group happened to center around the Roman Catholic Church and merely provided these individuals with a convenient substitute for the more ordinary lines of endeavor. (d) Whether such traits would occur as frequently among more mature members of the same profession, i.e., among the same individuals five or ten years hence or within a random sample of adult priests, it is impossible to say. Any generalization from the present data, based on younger men about to enter the priesthood, to members of the profession at all ages, is entirely a matter for conjecture. Social pressures, maturity, etc., might operate to minimize the traits noted at the outset of the religious career, however valid the description of motivation with respect to the initial choice of the profession. (e) No attempt is made to judge the desirability of religious adjustment, even though the observed temperamental differences were perfectly valid and represented more

or less fixed traits of behavior. "Adjustment" may imply either what society happens to regard at the present time as the "normal" or expected thing, or, psychologically, what is most efficacious and best adapted to the religious temperament as we actually find it in society.

SUMMARY

Temperament, as measured by self-ratings on the Heidbreder scales for introversion and the inferiority complex, is clearly related to the variety of religious experience which is evidenced by enrollment in a Roman Catholic seminary. The chief results of the study are as follows:

1. Pronounced differences in central tendency are found between the ratings of theology students and the ratings of college students, university faculty members, and business men on the scale for introversion, indicating greater frequency of the trait among divinity students.

2. Individuals enrolled in the seminary are more subject to inferiority attitudes than are people in general. The differences in central tendency are uniformly in the same direction, and are, with one exception, statistically significant.

3. Inferiority attitudes are not as common among advanced theology students as among younger men preparing for the priesthood. College freshmen show approximately the same central tendency in the ratings on inferiority attitudes as do advanced theology students, who are seven or eight years older on the average.

4. Correlations between ratings on introversion and inferiority attitudes and scores on the Woodworth Psychoneurotic Inventory are in the neighborhood of $+.50$, indicating that the traits measured by the respective scales are not discrete independent entities, but represent general emotionality rather than specific unrelated patterns.

5. Individually diagnostic items on both scales, determined by an item analysis of the ratings of theology students and college upperclassmen, corroborate the findings based on central tendency. The detailed pattern of traits characterizing seminary students may definitely be classed on the whole as symptoms of introversion and the inferiority complex as such. These traits do not represent merely stereotyped attitudes which are traditionally endorsed by the Catholic religion.

6. These results confirm the classic testimony of historic religious personalities described by James, Leuba, and other students of the subject.

REFERENCES

1. CAMPBELL, K. J. The application of extroversion-introversion tests to the insane. *J. Abn. & Soc. Psychol.*, 1929, **23**, 479-481.
2. CONKLIN, E. S. The psychology of religious adjustment. New York: Macmillan, 1929. Pp. xiv+340.
3. DESANCTIS, S. Religious conversion. (Trans. by H. Augur.) New York: Harcourt, Brace, 1927. Pp. 324.
4. FATERSON, H. F. Some implications of the normal inferiority complex. Ph.D. thesis, Univ. Minn., 1928.
5. GARRETT, H. E. Statistics in psychology and education. New York: Longmans, Green, 1926. Pp. 330.
6. FREYD, M. Introverts and extroverts. *Psychol. Rev.*, 1924, **31**, 74-87.
7. HEIDBREDER, E. Measuring introversion and extroversion. *J. Abn. & Soc. Psychol.*, 1926-1927, **21**, 120-134.
8. ———. Introversion and extroversion in men and women. *J. Abn. & Soc. Psychol.*, 1927-1928, **22**, 52-61.
9. ———. The normal inferiority complex. *J. Abn. & Soc. Psychol.*, 1927-1928, **22**, 243-258.
10. JAMES, W. The varieties of religious experience. New York: Longmans, Green, 1902. Pp. xii+534.
11. LEUBA, J. H. The psychology of religious mysticism. New York: Harcourt, Brace, 1925. Pp. xii+336.
12. SWARD, K. An experimental study of leadership. Ph.D. thesis, Univ. Minn., 1929.

Western Reserve University
Cleveland, Ohio

LE TEMPÉRAMENT ET L'EXPÉRIENCE RELIGIEUSE

(Résumé)

Le tempérament, mesuré par des auto-appréciations sur l'introversion et le complexe d'infériorité (échelles Heidebreder) a un rapport défini avec l'action de devenir étudiant dans un séminaire catholique. Le groupe de critère se compose de 80 étudiants du séminaire. Les normes comparatives sont basées sur les appréciations d'environ 2000 étudiants universitaires, de 52 hommes d'affaires et professionnels, et de 68 membres de la faculté de l'Université de Minnesota. On a résumé les résultats des études antérieures sur la constance et la valeur des deux échelles.

Des analyses statistiques des tendances centrales des données montrent que la soi-disant introversion se trouve beaucoup plus fréquemment chez les étudiants de théologie. Les résultats sur les attitudes d'infériorité, quoique moins frappants, montrent la même tendance. L'analyse des appréciations, fait pour fait, montre que le fait marqué de telles attitudes chez ceux étudiant pour devenir prêtres peut s'attribuer au tempérament en lui-même plutôt qu'à l'influence des standards stéréotypés en vogue parmi les sectes chrétiennes.

Si ces résultats représentent de vraies différences de comportement entre

les groupes considérés, on peut regarder de telles attitudes émotives comme des facteurs prédisposants et motivants pour la forme de l'expérience religieuse en question. James et Leuba sont arrivés à des vues quelque peu semblables d'une autre source d'évidence, c'est-à-dire, des biographies des saints religieux et des mystiques.

SWARD.

TEMPERAMENT UND RELIGÖSE ERFAHRUNG

(Referat)

Temperament, mittels Selbstabschätzungen in Bezug auf Introversion und Minderwertigkeitsgefühl (Skalen von Heidbreder) gemessen, steht mit der Tatsache des Eingeschriebenseins in einem römisch-katholischen Seminar in bestimmtem Zusammenhang. Die massgebende (criterion) Gruppe bestand aus 80 Studenten der Theologie. Vergleichswerte (comparative norms) gründen sich auf Abschätzungen von ungefähr 2000 Universitätsstudenten, 52 Geschäftsmännern und Fachmännern, und 68 Mitgliedern der Fakultät der Universität von Minnesota. Die Befunde über die Zuverlässigkeit (reliability) und Triftigkeit (validity) der zwei Skalen werden aus früheren Untersuchungen kurz zusammengefasst.

Statistische Analysen der Zentralrichtungen (central tendencies) der Befunde erweisen, dass die sogenannte Introversion unter Studenten der Theologie bedeutend häufiger ist. Die Befunde in Bezug auf Minderwertigkeitsgefühle neigen in die selbe Richtung, obwohl sie weniger auffallen. Analysiert man die Abschätzungen Punkt für Punkt, so wird erwiesen, dass die auffallende Häufigkeit solcher Einstellungen bei Kandidaten für das Priesteramt eher dem Temperament als solchem wie dem Einfluss von stereotypen Richtmassen (standards) welche bei christlichen Sekten herrschen, zuzuschreiben ist.

Stellen diese Befunde tatsächliche Unterschiede im Benehmen (behavioristic differences) zwischen den betreffenden Gruppen dar, so darf man solche Gemüteseinstellungen (emotional attitudes) als Faktoren betrachten, welche die betreffende Form der religiösen Erfahrung vorbereiten oder motivieren. James und Leuba gelangten mittels einer selbständigen Beweisquelle, namentlich der Biographien religiöser Heiliger und Mystiker, zu etwas ähnlichen Resultaten.

SWARD

SHORT ARTICLES AND NOTES

THE COLOR PREFERENCES OF JAPANESE CHILDREN*

THOMAS R. GARTH, KUNIHEI IKEDA, AND ROY M. LANGDON

A study of the color preference of Japanese children was reported by Megumu Imada in the *Japanese Journal of Psychology* of August, 1926 (6). The subjects of the experiment were 1059 school children from the first grade through the tenth grade. According to the writer of the article, the experiment was incited by the work of Garth on color preference; but the procedure, the color stimuli, and the method of handling the data were different. For this reason we have thought it wise to use Garth's method and procedure on a group of 1011 Japanese school children. This is in order to complete a series of studies in racial differences in color preferences made by Garth and his students on Indians (1), Whites (2), Negroes (7), and Mexicans (4), and again in order to bring all the studies into relation.

It is not deemed wise to take up valuable space in describing the afore-said experiment as it has been described in the references given above. It is sufficient to say that the colors were the Milton Bradley standard colors, red, blue, green, violet, orange, yellow, and white, all on coated papers. The method was the order of merit method from which color preference scales were derived, these being similar to the quality scales obtained by Thorndike (8) in handwriting and by Hillegas (5) in English composition. Not only were scales derived, but central tendencies were obtained for the various colors. (See Table 1.) However, the authors do not believe that the latter method is the better one for handling such data since the facts are not as clearly represented as by the scale method.

With these remarks we shall proceed briefly to give the results of the study made by the writers on the color preferences of Japanese. This experiment has been briefly reported by the senior collaborator (3).

SUBJECTS OF THE EXPERIMENT

The present study is from data obtained from Japanese school children living in the northern part of the island of Hondo by Mr. Kunihei Ikeda (M.A.) personally. The subjects were children from the first through the tenth grades. In the first seven grades both boys and girls were subjects, but in the eighth, ninth, and tenth grades only boys participated in the experiment. The ages ranged from 6 to 18 years. Ikeda reports that all

*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office, September 17, 1930.

TABLE 1
COLOR PREFERENCE SCALE FOR 1011 JAPANESE CHILDREN

Color	Scale value
White	.00*
Orange	.30
Yellow	.60
Green	.86
Violet	.86
Blue	1.43
Red	1.53

*White taken as a point of origin.

TABLE 2
COLOR PREFERENCE SCALES FOR 464 JAPANESE BOYS AND 420 JAPANESE GIRLS

Boys		Girls	
Color	Scale value	Color	Scale value
White	.00*	White	.00*
Orange	.34	Yellow	.49
Green	.95	Orange	.56
Violet	1.02	Green	.97
Yellow	1.06	Violet	1.01
Red	1.36	Blue	1.46
Blue	1.58	Red	1.91

*White taken as a point of origin.

TABLE 3
COLOR PREFERENCE SCALES FOR THE VARIOUS GRADES

First grade (N=125)		Second grade (N=130)		Third grade (N=130)		Fourth grade (N=125)	
White	.00	White	.00	White	.00	White	.00
Green	.78	Yellow	1.36	Orange	.95	Orange	.15
Orange	.93	Orange	1.58	Yellow	1.36	Violet	.76
Yellow	.97	Green	1.88	Green	1.40	Green	.83
Blue	1.19	Red	2.10	Violet	1.85	Yellow	.98
Violet	1.26	Blue	2.44	Blue	2.42	Blue	1.43
Red	1.71	Violet	2.51	Red	2.95	Red	1.73

Fifth grade (N=125)		Sixth grade (N=125)		Seventh grade (N=129)		Eighth, ninth, and tenth grades (boys) (N=127)	
White	.00	White	.00	Orange	.00	Orange	.00
Orange	1.25	Orange	.15	Yellow	.49	Yellow	.57
Yellow	1.29	Violet	.64	Violet	.56	Violet	.76
Green	1.74	Yellow	.90	Green	.94	White	1.02
Violet	1.74	Green	1.31	Blue	1.59	Green	1.17
Red	2.15	Red	1.65	Red	1.85	Blue	1.55
Blue	2.30	Blue	1.87	White	2.38	Red	1.70

the children tested were apparently full-blood Japanese, that they are from a district which is "off from the beaten track of civilization" and which is one of the most backward sections of Japan. Since most of the subjects could not write English, the experimenter was compelled to transcribe the records made by the subjects into English. While it is barely possible that there may have been some trace of Ainu blood in the subjects, the mixture must have occurred over a thousand years ago, since there are no Ainus to be found in this section of Japan at this time.

JAPANESE COLOR PREFERENCE SCALES

We give in Table 1 a color preference scale for 1011 Japanese children, in Table 2 color preference scales for boys and girls, and in Table 3 color preference scales for the several grades.

An examination of the color preference scale for the 1011 subjects shows the sequence for the colors to be, reading from most preferred to least pre-

TABLE 4
SHOWING AVERAGE AND MEDIAN POSITION ASSIGNED TO THE COLORS

Color	1011 boys and girls		Med.	Rank Med.
	Av.	S.D.		
Red	3.50	1.94	2.96	1
Blue	3.49	1.71	3.17	2
Violet	4.48	1.89	4.38	3
Green	4.45	1.74	4.52	4
Orange	5.31	1.71	5.54	6
Yellow	4.77	1.72	4.86	5
White	5.50	2.26	5.20	7

Color	464 boys (seven grades)		Med.	Rank Med.
	Av.	S.D.		
Red	3.83	2.03	3.48	2
Blue	3.32	1.69	2.96	1
Violet	4.44	1.94	4.31	3
Green	4.54	1.75	4.63	5
Orange	5.37	1.67	5.60	6
Yellow	4.32	1.65	4.31	4
White	5.67	2.16	6.64	7

Color	420 girls (seven grades)		Med.	Rank Med.
	Av.	S.D.		
Red	3.03	1.67	2.52	1
Blue	3.62	1.70	3.29	2
Violet	4.44	1.84	4.36	3
Green	4.50	1.69	4.57	4
Orange	5.08	1.76	5.26	5
Yellow	5.15	1.72	5.41	6
White	5.68	2.22	6.75	7

ferred: red, blue, violet, green, yellow, orange, and white. The colors violet and green have the same value, and none of the adjacent colors are seen to be very far apart on the scale. Moreover, the scale is rather short since the most preferred color, red, has a scale value of 1.53. This would indicate some confusion on the part of the subjects to discriminate, in the light of scale interpretation, decided color preferences. The chances are that another group might differ somewhat in their color preferences from this group. And this is what we find when we take Imada's results in comparison with this sequence, for he found the sequence to be blue, red, green, yellow, violet, and orange. He did not use white in his experiment. However, the two studies do agree in that orange is placed low, and red and blue relatively high. To be sure, the color violet is below yellow and green in his study, while it is equal to green in this study. Imada used central tendencies instead of obtaining scale values. However, when the writers of the present study found the medians, these corresponded in ranks to the scale sequence for the several colors, except that violet is preferred to green by a small margin.

If we take the scale value of orange as a base, we find that red "pulls" five times as hard, and blue almost as much, as does orange. Yellow "pulls" twice as hard, and violet and green nearly three times as hard, as orange.

SEX DIFFERENCES

If we may judge by the relative length of the scales, the girls would appear to have a somewhat better discriminative feeling for color than the boys. The girls prefer red most of all and the boys blue. There are other points of disagreement between the boys and girls, as when the girls place yellow all but last and the boys place it third from best. The boys seem to value yellow and violet about equally, while the girls find a decided preference for violet above yellow. Further observations may be made.

CHANGES ACCOMPANYING SCHOOLING

It is not claimed that education in the schoolroom is the only influence causing changes in color preference but it is interesting to note that while white remains the point of origin up through the sixth grade it rises to first place in the seventh grade. Since the eighth, ninth, and tenth grades are composed only of boys, it would not be safe to make observations there on that account. It is interesting to note that green rises to fourth place with the second grade and maintains that position, except in the sixth grade where it attains fifth place. Red is displaced by blue in the fifth and sixth grades. Violet, after some fluctuation, tends to fall as we proceed from grade to grade. Yellow, after some fluctuation, tends to fall. It is

interesting to note the shift of position that many of the colors experience from grade to grade.

Unfortunately, the data do not seem to permit of the crucial test of getting measured correspondence by the method of correlation between schooling and position of colors.

But it is seen to be a fact that colors shift their positions when we observe from this angle. This is most evident when we take the sequence of the first grade and compare it with that of the seventh grade.

RACIAL COLOR PREFERENCE SCALES

For purposes of comparing the Japanese with other races we give, in Table 5, color preference scales for various races which have been obtained.

TABLE 5
COLOR PREFERENCE SCALES FOR VARIOUS RACES

Race	No.	White	Yellow	Orange	Violet	Green	Red	Blue
Japanese	1011	0	60	30	86	86	154	143
Whites	1000	0	69	100	107	126	122	204
Indians	1014	0	65	110	189	144	238	174*
Negroes	1006	0	61	92	87	87	72	168
Mexicans	1152	7	0	14	21	70	74	55

*Garth's original scale for Indians was derived from a group of 559 full-bloods, but we report here a scale obtained from a group of 1014 full-blood Indians.

It will be seen that white is the common point of origin of all the scales for the races except in the case of the Mexicans where yellow is the point of origin. But in this case white is so little better than yellow that these colors may be said to have practically the same value. For purposes of comparison we may for convenience take white as a common point of origin for all races. At least we can make these observations: The Japanese scale being shorter than that of the Indians, the Whites, or the Negroes, these subjects do not discriminate preference for colors as clearly as do these racial groups. But they have a stronger feeling of difference for these colors than do Mexicans. The Japanese, Indians, and Mexicans agree in placing red first.

But the writers cannot feel justified in calling these racial differences in color preference for the reason that the factors of change due to different tribal notions and practices and other nultural influences can be shown to operate to modify the color preferences of the young for all races as we have shown in the case of the Japanese. In order to get at racial differences in color preference we must catch the young racial subject before these factors of change have operated.

REFERENCES

1. GARTH, T. R. Color preferences of five hundred and fifty-nine full-blood Indians. *J. Exper. Psychol.*, 1922, **5**, 392-418.
2. ———. A color preference scale for one thousand white children. *J. Exper. Psychol.*, 1924, **7**, 233-241.
3. ———. Race psychology. New York: Whittlesey-House (McGraw-Hill), 1931. Pp. xiv+260.
4. GESCHE, I. The color preferences of one thousand one hundred and fifty-two Mexican children. *J. Comp. Psychol.*, 1927, **7**, 297-311.
5. HILLEGAS, M. B. A scale for the measurement of quality in English composition by young people. *Teach. Coll. Rec.*, 1912, **13**.
6. IMADA, M. Color preferences of school children. *Jap. J. Psychol.*, 1926, **1**, 373-393.
7. MERCER, F. M. The color preferences of one thousand and six negroes. *J. Comp. Psychol.*, 1925, **5**, 109-146.
8. THORNDIKE, E. L. Handwriting. *Teach. Coll. Rec.*, 1910, **11**, 1-81.

University of Denver
Denver, Colorado

A COMPARATIVE STUDY OF THE INTELLIGENCE OF CHINESE CHILDREN ON THE PINTNER PERFORMANCE AND THE BINET TESTS

C. W. LUH AND T. M. WU

HISTORICAL COMMENT

Intelligence studies of the children of the Chinese immigrants to America and to Hawaii have recently been summarized by Hsiao (4). By way of historical orientation the present writers merely wish to add certain interpretations not emphasized in that article.

With respect to verbal intelligence the Chinese immigrant children seem to be much inferior to native whites, but non-verbal tests do not indicate much of a difference one way or the other. This language difficulty has often been explicitly recognized, but the full significance of the problem has just as often been overlooked. Language handicap exists where it is not even suspected. Graham, for instance, upon discovering that there is a marked difference between the Americans and the Chinese in auditory span for digits, remarks that "as the English names for figures are among the earliest things learned by the Chinese in the American school it is not probable that language difficulties disturb this test" (3, p. 50). We like to report here that Chinese college students who may speak English quite fluently have, on the average, an English memory span one to two digits shorter than their Chinese memory span.¹

¹A similar difference in auditory memory span for digits exists in German-speaking American college students, a problem worthwhile investigating.

Furthermore, the problem is often too lightly and perfunctorily dispensed with. Referring to Graham again, for example, differences in reading ability are supposed to have been eliminated by "holding constant" scores made on the Thorndike-McCall reading test. Neither the test nor the technique of partial correlation claims to do such wonders.

We criticize Graham, not because her work is particularly in error, but rather because it is one of the most exhaustive attempts that have been made. We believe that in such differential studies no existing verbal test can be an adequate tool.

Returning now to non-verbal tests, it seems to us that Sandiford and Kerr's work (5) on the children of British Columbia with the Pintner-Paterson performance tests has been the most significant. Our own attempt is nothing but a continuation of their effort. According to their estimate, the average performance IQ of 226 Chinese children is 107.4. Their conclusions and interpretations are based upon this and similar figures which represent the children's standing on the so-called percentile scale. Evidently they have overlooked the inconsistency in their own and also Pintner's original data between the percentile scale, on one hand, and the point scale and the median mental age scale, on the other, the last two being mutually consistent, more or less. Even without making allowance for retardation in school grade which an immigrant group will almost inevitably suffer, the average IQ of the Chinese children tested by Sandiford and Kerr could not have been higher than 100 according to the two latter scales.

While this discrepancy in interpretation is not an inherent difficulty for testing work of this kind, there are other handicaps which even a non-verbal test cannot avoid. First of all, we cannot estimate how much of a selective factor Chinese immigration is. Opinions in China as to the mental quality of the Chinese emigrants to America run to opposite extremes. In addition, a non-verbal test may also have its informational difficulties when its contents become very concrete, such as Healy Picture Completion I in the Pintner series. And we must not forget the influence of the notorious "white tester" among "colored" groups.

THE PURPOSE OF THE PRESENT STUDY

As already stated we merely wish to continue the work of Sandiford and Kerr. We try to apply the most widely used individual performance tests to a sample of Chinese children in China under influences which can be described and estimated. We cannot depend upon random sampling, an impossible step under China's present educational situation. On the other hand, we give the children at the same time Luh's revision of the Binet tests. The two scales being not highly correlated, the Binet data can give us only a rough gauge of the status of the children. However, it would be more difficult to grope in the dark.

Incidentally, we also try to see how children in China respond to the con-

tents of the Pintner-Paterson tests which are in certain respects typically American.

THE CHILDREN TESTED

The number of children tested is 128, of which 106 attend the first seven grades of the elementary school of the Department of Education of Yenching University. Of these 106, 37 belong to homes of professorial and similar standing. These children are designated as Group A in this study. Thirty-five are children of the clerical class, and compose Group B. The other 34, who form Group C, are children of day laborers and household servants. The last group is composed mainly of natives of the local district, while the first group comes from all over the country. Thus we have a composite group very much like the one Chase and Carpenter (1) or Dashiell and Glenn (2) worked on. The remaining 22 children are taken at random in Peiping and the Yenching vicinity to serve as a further check on the larger group. They are to be known as Group D.

In age distribution the groups differ. The Binet and the performance tests are not given on the same day to any single child, so the age averages, together with their *S.D.*'s, are given separately for the two scales in Table 1.

There is a tendency among all these groups for age to be negatively correlated with performance IQ. In the order of decreasing magnitude, the coefficients range as shown in Table 2.

But with the Binet tests age is not correlated with IQ at all, except in the

TABLE 1

Group	Age when given Binet test	<i>S.D.</i>	Age when given Pintner test	<i>S.D.</i>
Group A	9.2	2.2	9.1	2.3
Group B	9.6	2.5	9.6	2.5
Group C	11.6	2.7	11.4	2.7
All Yenching children	10.1	2.7	10.0	2.7
Group D	10.9	2.8	10.9	2.7
All children tested	10.2	2.7	10.2	2.7

TABLE 2
CORRELATIONS BETWEEN AGE AND PERFORMANCE IQ

Group	<i>r</i>	<i>P.E.</i> _r
Group A	— .43	.09
Group B	— .36	.10
Group C	— .35	.10
All Yenching children	— .45	.05
Group D	— .28	.13
All children tested	— .43	.05

case of Group C which has a coefficient of $-.38 \pm .10$. One might thus conclude that for the Binet tests the children are not selected with reference to age as far as their own social standing is concerned. With the performance tests, however, all the groups become thus selected, which shows that the Pintner norms really cannot be applied in the present case. Either the norms for the younger ages are too easy, or those for the older ages are too difficult, or both are at fault. Judging from scattered American data and general observations in the present study, we rather think that the second alternative has been the most effective. Without further investigation we have to take our own data for what they are worth. We hope our comparative interpretations are not thus affected.

THE CHINESE BINET SCALE AND ITS RELATION TO THE PINTNER MEDIAN MENTAL AGE SCALE

The Chinese revision of the Binet tests as standardized in 1924 is of the nature of a point scale with $1/10$ of the *S.D.* of the 12-13-year-old distribution as the basic unit. In content it is like Herring's revision. No detailed comment is needed here except that the standardization was carried out under very unfavorable conditions and consequently it has been generally felt that the norms are inadequate and somewhat too low. While we are not in a position to make definite corrections, it is safe to say that with a group averaging around 110 IQ points as ours does probably a discount of 5 points should be made. This we shall remember in our comparative statements.

As a rule a child takes these tests several days after the performance tests have been given. In a few exceptional cases the intervals are as long as 10 months.

The correlation coefficients, shown in Table 3, with age "partialled out," can give us some idea of the extent of relationship between the two scales.

It will be seen that with Group C the two scales are practically uncorrelated. Group B agrees with Group D in giving the highest correlation coefficient. The two groups are quite similar, as one can see from the data in the next section. We think the two scales are correlated at least to that extent. Group B, in fact, is more representative of a general

TABLE 3

Group	<i>r</i>
Group A	.41
Group B	.56
Group C	.12
All Yenching children	.49
Group D	.54
All children tested	.49

TABLE 4

	Average Binet IQ	<i>S.D.</i>	Average Pintner IQ	<i>S.D.</i>
Group A	130	22	117	23
Group B	112	26	103	18
Group C	98	17	94	12
All Yenching children	114	26	105	21
Group D	113	22	101	21
All children tested	114	25	105	21

city population than either the whole Yenching group or all the children tested. We have already described the Yenching group as a composite group with the two ends of the distribution selected, especially the lower one.

THE RESULTS OF THE BINET AND THE PINTNER TESTS

We shall first present the average IQ's for the separate groups on the two scales (Table 4).

The above tabulation needs a little amplification. As to the Binet averages, we have to remember the amount of correction that must be made. The averages for Group B and Group D are 112 and 113, respectively; these agree very well with that for all the Yenching children or all the children tested. Our estimate is that the group averages around 108 or 109, i.e., 5 points below the apparent figures. We cannot be far away from the truth.

It is, however, the performance averages that require closer examination. The general average of 105 is 4 points higher than that for Group D which is the "random" group. Referring back now to the tendency for age to be negatively correlated with Pintner IQ, we find it is in this small group that the coefficient is hardly significant. The group contains a smaller percentage of very young children than either Group B or all the Yenching children. If we eliminate from the groups all children below age 8, the averages become 100.0 for Group B, 100.6 for Group D, 100.4 for all Yenching children or all the children tested. The Pintner norms for the younger ages are again to be questioned.

Our main point is next to compare the averages from the two test scales. Are we to accept the performance averages at their face value? To answer this question we have to report some supplementary data.

1. According to the median mental age scale no child can have a mental age higher than 14. Of the 128 children tested 10 actually go beyond the limit. Their IQ's could be somewhat higher if the scale did not stop at that limit. However, it is not very probable that the average for the whole group could be much raised.

TABLE 5

	For 22 items	For 21 items	For picture completion only
Group A	117	117	107
Group B	103	105	89
Group C	94	97	71
All Yenching children	105	107	89
Group D	101	103	80
All children tested	105	106	88

2. Many of the children have difficulty with Healy Picture Completion I. For instance, very few Chinese children have ever seen American football, it is most alluring to use in its stead the big pumpkin or the baseball as many of them do. Quite a few children put the stool where the piece of log ought to belong, also a very natural scene. So let us eliminate the item for picture completion from the 22 separate mental age estimates for each child and then take the median mental age and calculate the IQ. The averages thus obtained are slightly higher, though the brighter children are not affected.

3. The Pintner scale being a composite one, some of the test elements have norms only for a very limited age range. The mannikin test and the adaptation board test can be used only up to age 8, the feature profile test only down to age 12. Evidently these particular tests are unfitted for a group with as large an age variation, as the one we have tested. If we eliminate these three items in the same way as we do the picture completion test and take median mental ages and then IQ's from the remaining 11 tests or 18 items, the averages are raised as follows:

Group A	from 117 to 121	
Group B	103	105
Group C	94	98
All Yenching children	105	108
Group D	101	105
All children tested	105	108

We are now in a position to make a comparative statement. This group of 128 Chinese children who fairly represent the general urban population has an average Chinese Binet IQ of about 108-109. With proper treatment the Pintner performance IQ for the same group also averages around 108. The two test scales are correlated to the extent of .50 or more with age "partialled out." Roughly speaking, therefore, the Chinese children have about the same degree of performance intelligence as American children, insofar as the Pintner norms are adequate for the general American population.

REFERENCES

1. CHASE, H. W., & CARPENTER, C. C. The response of a composite group to the Stanford revision of the Binet-Simon tests. *J. Educ. Psychol.*, 1919, **11**, 179-189.
2. DASHIELL, J. F., & GLENN, W. D. A re-examination of a socially composite group with the Binet and with performance tests. *J. Educ. Psychol.*, 1925, **16**, 335-340.
3. GRAHAM, V. T. Intelligence of Chinese children in San Francisco. *J. Comp. Psychol.*, 1926, **6**, 43-72.
4. HSIAO, H. H. The mentality of the Chinese and Japanese. *J. Appl. Psychol.*, 1929, **13**, 9-31.
5. SANDIFORD, J., & KERR, R. Intelligence of Chinese and Japanese children. *J. Educ. Psychol.*, 1926, **17**, 361-367.

Yenching University
Peiping, China

BOOKS

J. W. BRIDGES. *Psychology, normal and abnormal*. New York: Appleton, 1930. Pp. xxii+552. \$3.50.

"This book presents the main facts and theories of normal and abnormal psychology and attempts to relate them to the actual and assumed functions of the nervous system. It sets forth at the same time a system of psychology. The standpoint is similar to that defined by Warren in his *Human Psychology*. It is based upon the double-aspect theory of mind-body relation, and thus recognizes both introspective and behavioristic methods. Psychology treats of both consciousness and behavior. The problem of the relationship of these two aspects of experience is squarely faced from the beginning; and it is held that a philosophical foundation is essential for psychology, as indeed it is for all sciences.

"The point of view adopted permits of eclecticism in the selection of material and often leads to reconciliation of seemingly divergent views. The text, therefore, includes contributions from the various schools of psychology. It comprises important findings and significant doctrines of structural and functional psychology, of behaviorism and psychoanalysis.

"The book was written with special reference to the needs of medical students and practitioners, . . . no suitable textbook is yet available for a course for medical students. The writer has attempted to meet the requirements of such a course. He feels that the general standpoint of the book as well as the following features will appeal to the medical profession.

"First, psychological facts and theories are intimately linked with those of physiology. . . .

"Secondly, abnormal mental processes and behavior are described in connection with the normal, but the book does not presuppose any knowledge of normal psychology. . . .

"Thirdly, the attempt has been made to define and delimit the field of psychology."

The following points arrest our attention in the above quotation:

1. The book attempts to give a systematic account of normal and abnormal psychology.
2. It proposes to set forth a system of psychology and remain eclectic with reference to choice of material.
3. It asserts that a philosophical foundation is essential for psychology.
4. The standpoint is based upon the double-aspect theory of mind-body relation. The inference is made that no other standpoint makes possible an all-inclusive consideration of psychology.
5. It recognizes both existential and behavioristic methods.
6. "Psychological facts and theories are intimately linked with those of physiology."

7. "Abnormal mental processes and behavior are described in connection with the normal."
8. An attempt is made to define and delimit the field of psychology.

We shall now consider these points in the above order:

1. In harmony with the plan followed in his *Outlines of Abnormal Psychology*, the author treats first the normal, then the abnormal aspects of the various psychological phenomena.

Psychology is defined in terms of the stages in the evolution of the science, such as the study of: the soul, the mind, consciousness, behavior, and behavior and consciousness. The author takes the latter of these views, namely, that psychology is the science of both consciousness and behavior. Consciousness is analyzed into major parts and also into finer units or elements. These conscious elements are then discussed from a structural and functional point of view. The major aspects of consciousness are: (1) cognitive (the knowing aspect of consciousness), including perception, memory, imagination, reasoning, and intellect; (2) affection (the feeling aspect), including feelings of pleasantness and unpleasantness, and also emotions, moods, and temperaments; and (3) conation (the striving and doing aspect), including instinctive and habitual impulses or drives, volition, inhibition or control, and character.

The elements of consciousness are: (1) sensation, (2) feeling, and (3) impulse. "Most psychologists agree that sensation is the chief element of cognitive consciousness." The primary characteristics or attributes of sensation are four in number, namely (1) quality, (2) intensity, (3) clearness, and (4) duration.

When consciousness is considered as a process the following aspects are outstanding: (1) It is a "stream of consciousness" manifesting a succession of "flights and perchings." "This stream of consciousness is related to, and identical with, the stream of neural activity." (2) It possesses "personal identity," i.e., a temporal continuity. (3) It is a "unit" or a "complete whole." (4) It possesses as its essential characteristic associative powers. The phenomena that become associated in consciousness are: ideas, sensations, perceptions, impulses, feelings and emotions, and the effect of a stimulus upon a sense-organ, namely, the "conditioned reflex."

The behavior side of Dr. Bridges' psycho-biological duality is most clearly characterized by reflexes, instincts, human impulses, and habits. But each of these has also subjective or conscious aspects. "Reflexes," for example, "differ considerably with reference to their conscious aspects. The simplest reflexes, such as the pupil reflex and the various tendon reflexes, have little or no conscious aspect but the more complex reflexes are usually experienced by the subject as conscious impulses or drives. . . . An instinct may be objectively a *pattern reaction*, and at the same time subjectively a fundamental *impulse* or *drive*. As a behavior pattern it is determined by a neural pattern, and as an impulse, it is the subject of innervation of the same neural pattern. . . . From the objective point of view habit may take two forms. First,

it may be conditioned response analogous to the "conditioned reflex. . . . The second form of habit is a coordination of response. . . . From the subjective point of view, habits like instincts are impulses. They are either impulses conditioned to new situations, or associations of simple impulses into new combinations. The formation of habit is thus the organization of impulses into systems or hierarchies." A similar duality is observed in memory, attention, perception, and emotions, whereas ideas, images, and feelings are primarily subjective experiences.

2. Dr. Bridges attempts to set forth a system of psychology and remain eclectic with reference to choice of material. Eclecticism versus system making has been recently discussed by Klein (2), who raises the question as to "why we have systems of psychology but no systems of chemistry, physics, astronomy, or physiology?" He suggests, "May it be that this phenomenon is indicative of a still active philosophical trend among psychologists, . . . a trend that is anachronistic among the natural sciences?" (p. 488). In answer to this question, Brown (1) replies that there always has been and always will be system-making in the various natural sciences, including physics and psychology.

Regardless of the ultimatum that system-making will suffer at the hands of scientists, does not the adoption of a system invariably lead to a selection of those data and theories which fit into the given system? One wonders to what extent Dr. Bridges was able to remain eclectic in his choice of theories concerning the mind-body relationship, consciousness and behavior, the subjective and the objective treatment of sensations; and in the choice of material pertaining to perceptions, instincts, emotions, memory, etc.? It seems a rather ambitious undertaking to attempt straddling the issues of system-making and eclecticism at the same time.

3. Dr. Bridges' assertion that a philosophical standpoint is essential for psychology, is likewise subject to considerable difference of opinion. This is especially true if we mean by philosophical standpoint adherence to one of the several views concerning the mind-body relationship. In view of the diversity of opinion concerning the meaning of such terms as "mind," "consciousness," etc., a number of authors of recent psychology textbooks have ignored philosophical standpoints relating to a mind-body relationship.

4. The standpoint of this book is based upon the double-aspect theory of the mind-body relationship. Other theories concerning the mind-body relationship are criticized and it is pointed out that the double-aspect theory "is in many ways the most satisfactory working hypothesis for both psychologist and physician." According to the double-aspect theory, "mind and body are not two separate and distinct entities. There is only one process but it has two aspects. . . . It is frequently convenient and it may help toward directness of expression to refer to consciousness as an aspect of neural activity, but the latter may equally well be referred to as an aspect of consciousness." Strictly speaking, they are both aspects of some

more fundamental process that is perceived objectively as neural activity and is experienced subjectively as consciousness. The question now arises: What is this "more fundamental process" of which the two are aspects? Does Dr. Bridges mean that they are identical in the sense of both belonging to the "one process,"—if so, they are not strictly *identical* but are *two* aspects of an identical process, and this duality of aspects and their relation to each other must still be explained. Dr. Bridges thus fails to show how neural activity can give rise to anything but *responses*. If responses compose the behavior aspect of the proposed duality, what does there remain to bring about the conscious aspect?

5. Dr. Bridges states that his point of view "recognizes both introspective and behavioristic methods." This implies a fundamental difference of methodology for the various schools of psychology. However, by a mere examination of methodology as such, it is not at all obvious by which school a given method was contributed. Ebbinghaus' method of studying memory, for instance, would fit into the behaviorist as well as in the structuralist or existentialist program. Though systems have come and gone, the methods which they contributed have gained a permanent place in the evolution of the science. There is no further need of keeping psychology esoteric by a delimitation of its methodology and terminology.

6. "Psychological facts and theories are intimately linked up with those of Physiology." The author has religiously followed this dictum, and, consequently, is not guilty of Lashley's otherwise just criticism that: "The chapter on the nervous system seems to provide an excuse for pictures in an otherwise dry and monotonous text. That it has any other function is not clear; there may be cursory references to it in later chapters on instinct and habit, but where the problems of psychology become complex and interesting the nervous system is dispensed with" (3, p. 1-2).

7. "Abnormal mental processes and behavior are described in connection with the normal." Approximately 37 per cent of this book is devoted to abnormal phenomena. This includes a chapter on psychopathology. The point of view adopted for a consideration of abnormal phenomena is consistent with that of normal phenomena. The author feels that "abnormal psychology is without doubt the best approach to the understanding of human nature. Abnormal behavior and mental processes are merely exaggerations or diminutions of normal functions. Normal processes are thus thrown into relief and can more easily be studied."

8. In harmony with his views expressed elsewhere the author defines and delimits the field of psychology. "The writer feels that psychologists in the past have tended to encroach upon the practical fields of medicine, education, social work, and other professions, instead of merely contributing to the scientific foundations of work in the applied fields. It is shown that psychology has the same value for the medical man as anatomy and

physiology have, and that there is no more reason for conflict between psychology and medicine than between physiology and medicine."

REFERENCES

1. BROWN, J. F. A note on Dr. Klein's plan for eclecticism. *Psychol. Rev.*, 1931, **38**, 182-185.
2. KLEIN, D. B. Eclecticism versus system-making. *Psychol. Rev.*, 1930, **37**, 488-496.
3. LASHLEY, K. S. Basic neural mechanism in behavior. *Psychol. Rev.*, 1930, **37**, 1-24.

P. H. EWERT

University of Vermont
Burlington, Vermont

SOME NOTES ON THE CENSUS OF RELIGIOUS BODIES, 1926*

From the Department of Biology of the School of Hygiene and Public Health of The Johns Hopkins University

RAYMOND PEARL

I

In his Lowell Lectures, Whitehead (7) made the following statement:

"The present state of religion among the European races illustrates the statements which I have been making. The phenomena are mixed. There have been reactions and revivals. But on the whole, during many generations, there has been a gradual decay of religious influence in European civilization. Each revival touches a lower peak than its predecessor, and each period of slackness a lower depth. The average curve marks a steady fall in religious tone. In some countries the interest in religion is higher than in others. But in those countries where the interest is relatively high, it still falls as the generations pass. Religion is tending to degenerate into a decent formula wherewith to embellish a comfortable life" (7, p. 262).

That the trend which is envisaged by Whitehead over a long-time base is a matter of concern to those professionally interested in religion at the present moment is indicated by the following quotation from the *Wesleyan Pastoral* drawn up by direction of the 1928 Liverpool Conference, and signed by its president, the Rev. John W. Lightley. The quotation is taken from the text of the *Pastoral* published in the *London Times* of August 16, 1928.

"But it is surely a matter of grave concern that after all these years of evangelism, and through an era of church building unprecedented, when knowledge has increased, and the foundations of the Christian faith have been tested and proved, and with an organization so vast and perfected, the Christian Church should count so little in the ordinary life of the people. That is the disquieting element in the religious outlook today.

*Received in the Editorial Office, May 22, 1931.

In spite of all its activities, and its earnest efforts to attract and win the people, the Church is yet surrounded by an indifference which is sufficient to make the most complacent uneasy. The Gospel is not challenged so much as it is ignored. Paganism abounds in the land, and is constantly fed by new movements and interests—some of which present themselves in the guise of healthgiving, aesthetic, or spiritual tonics. The sense of the necessity for worship stirs the people little, and multitudes at the very doors of the Church are living as though Christ were dead."

The priests of older and more conforming faiths than the Methodists are noting a similar phenomenon. It has been for many years a custom of the Bishop of London to commission annually one of the clergy of his own diocese to write a book to be read by the flock during Lent. In 1930 this doubtless pleasant and certainly pious duty was assigned to the Rev. W. P. G. McCormick (3), the Vicar of St. Martins-in-the-Fields. The book is a plea for joy in the fellowship of the Church. But in it one finds the following revealing note:

"If we parsons don't count in the minds of ordinary men, it is not because our office is despised; it is because we present to the world a wrong view of that office. If we suffer—as well we may in an age of perpetual criticism of the Church, its ministers and all its doings—from an "inferiority complex" the average man is not going to help us to sublimate it. He is not enough interested in the Church or its officials to make the attempt; and we shall be taken at our own low valuation."

A strikingly similar melody, though in a more colorful key than the C major of the previous quotations, is intoned by a young American clergyman, who prefers anonymity (1). After a year of experience this eager youth of 25 is plainly discouraged. He finds it difficult to get and hold young people in his congregation. He says:

"No doubt I could have enticed them into the sheepfold with a ham sandwich and a song-and-dance act.... But the thought repulses me, for I dread the day when the Protestant church will be known as a place to get a good chicken dinner instead of a place to worship.... A year in the ministry has set me to wondering if deep down in the heart of youth there really is a desire to follow the young man Jesus."

This disillusioned parson has had revealed unto him that the prohibition of various things—by no means alcohol only—has become the keystone of Protestant thought and activity. He says:

"The Protestant church rests upon the wabbly foundation of the success it has achieved in reforming the world morally. It rests upon Sinai and not upon the Mount made sacred by the great sermon of Jesus. . . Why can't we say quite frankly that the great majority of moderns don't give a cuss about the church or Him it represents?"

There is no need to multiply further these quotations. They sufficiently indicate that there is presented here a problem of social psychology of great interest. Are the trends suggested by the quotations which have been given capable of definite statistical proof and measurement in the American population at the present time?

II

Regarding social psychology as a highly important branch of human biology in general, I (4) have ventured to suggest that:

"Any attempt to describe or objectively evaluate the psychology of a group, as such, is bound to be a difficult matter. And yet every one knows that there is such a thing as a societal psychology—a point of view of the group as group, a reaction pattern which characterizes the group and differentiates it from other groups. There are two important reasons why no one can expect general agreement to any conclusion he may draw regarding any particular aspects of societal psychology which have real interest. The first of these reasons is that any such question of more than purely academic interest is sure to involve a matter about which nearly everyone has emotions and prejudices all his own, which consciously or unconsciously color his views about anything said on the subject. The second reason is that societal psychology lacks a critical, time-seasoned, objective methodology."

Regarding methodology the further point was made that:

"Any sort of group biology, using the term biology in the broadest sense, must be dependent upon statistics. Statistics form the basic observational data for groups. It is by putting statistical data of different sorts, each in a different way pertinent to the subject under investigation, *into relation with each other*, that there lies the hope of building up a critically sound science of societal psychology, as it seems to me."

The primary difficulty, of course, is to find official statistics that have any pertinence regarding psychological problems. Only occasionally are they published. Fortunately, the Census Bureau has

lately been occupied with, and has now completed the publication of, an extremely useful and penetrating census of religious bodies in the United States, as of the year 1926 (6).

The following general statements, quoted from the bulletins already published on the census, will give an idea of the scope and technique of the work:

"The census of religious bodies, as its name indicates, is a census of religious organization rather than a census of population according to denominational affiliation. This census is taken once in 10 years in accordance with the provisions of the permanent census act approved March 6, 1902, and is confined to churches located in continental United States.

"Reports are obtained by the Bureau of the Census from each of the congregations, churches, or other local organizations of each religious body. The census data are thus obtained directly from the local churches and are not in any sense a compilation of the statistics collected by the different denominations and published in their yearbooks. Lists of the local organizations for 1926 have been secured in most cases, however, from the denominational headquarters, and much additional assistance has been rendered by the officials of the various denominational organizations.

"The statistics of expenditures relate either to the calendar year 1926 or to the church record year which closed at some time during 1926, and the figures for membership, value of property, etc., relate to the close of that year.

"*Number of churches.*—In the reports of the Census of Religious Bodies for 1916 and 1906 the total number of churches or organizations shown for some denominations was slightly in excess of the number of churches reporting membership. Since membership figures have been obtained for all of the churches included in the reports for the current year, and for other reasons, it has seemed advisable to use for purposes of comparison with 1926 the number of churches reporting membership in 1916 and 1906. These figures are used, therefore, in the tables representing comparative figures for earlier censuses.

"*Membership.*—The members of a local church organization, and thus of the denomination to which the church belongs, are those persons who are recognized as constituent parts of the organization. The exact definition of membership depends upon the constitution and practice of the church or denomination under consideration. Each church was instructed to report the number of its members according to the definition of mem-

ber used in that particular church or organization. In some religious bodies the term 'member' is applied only to communicants, while in others it includes all baptized persons, and in still other bodies it covers all enrolled persons."

III

It is the purpose of the present paper merely to discuss in an entirely preliminary way some of the more interesting information derivable from this census as a whole.

Let us consider first the *growth* of religious bodies in the United States in the decade from 1916 to 1926 in comparison with the growth of the total population of the country. The necessary data are given in Table 1, in which figures for Jewish congregations are omitted, as they are not comparable at the two censuses.

TABLE 1

GROWTH IN THE PERIOD 1916-1926 IN (a) NUMBER OF CHURCHES, (b) CHURCH MEMBERSHIP, AND (c) TOTAL POPULATION IN THE UNITED STATES

Year	Churches	Membership	Population*
1926	229,036	50,495,104	117,136,000
1916	225,099	42,954,512	100,757,735
Gain in absolute numbers	3,397	7,540,592	16,378,265
Percentage gain	1.7	17.6	16.3

*Official estimates.

At first and superficial examination the facts set forth in Table 1 might seem to indicate that the *Wesleyan Pastoral*, quoted above, takes an unduly pessimistic view, at least so far as concerns the United States. Church membership increased between 1916 and 1926 at slightly more than the rate of the growth of the population as a whole. To be sure the number of churches increased in the decade only 1.7%, but it is an entirely reasonable point of view, at least economically, to regard this as a sound development. Relatively fewer churches took care of relatively more people in 1926 than in 1916. But it is to be noted that, whereas in 1916, 42.6% of the total population were church members, in 1926 the percentage had risen to 43.1% (exclusive of the Jewish in both cases). This cannot be regarded as an especially vigorous growth.

The social functions of organized religion are diverse, but no one would deny that it is a significant element in human societal organization. In earlier times it embodied almost the only, certainly the

most extensively, organized social diversion from the ordinary affairs of workaday life. It has always traditionally been held to be a moral element in the structure of society, making for stability and conservatism.

Now it is a fair thing, and an important one from the standpoint of social psychology, to inquire how the growth of organized religion, in the decade under consideration, compares with the growth of some other elements in the societal structure which have social functions at least analogous to those which it is supposed to perform. To enable the formation of a rough, approximate judgment on this point, Table 2 has been prepared, the data being drawn entirely from official sources (principally the *Statistical Abstract* of the United States, *passim*). Unfortunately, data are not available in all cases for the exact years 1916 and 1926. I have indicated, however, in each case the years actually taken, and have chosen from those available the ones coming nearest to the two mentioned, except in a few cases where for obvious reasons to do so would lead to incorrect or exaggerated deductions.

National banks may perhaps fairly be regarded as stabilizing, conservative elements in the societal structure as a whole. From Table 2 it first appears that the number of such banks in the country increased at rather more than three times the rate that the churches did in the decade from 1916 to 1926.

The deposition of money in a savings bank is generally regarded as a meritorious and conservative sort of action, socially analogous in some degree certainly to church membership. But the number of such depositors increased relatively more than twice as much as did the number of church members between 1916 and 1926.

According to one's tastes, there are other ways available for gaining diversion from the ordinary tasks of life, besides being a church member. Data relative to some of these other ways are assembled in the portion of Table 2 lying to the right of the first three columns. It appears, first, that the consumption of alcoholic beverages increased by 4.5% in the single year of 1916 to 1917. The consumption of such beverages *per capita* of the *whole* population increased in that year by more than half a gallon, from 19.61 to 20.20 gallons per head. The year 1917 is the last that can be fairly used in this particular comparison, because war-time prohibition on a nation-wide scale came in during 1918. After that date the whole population

TABLE 2
GROWTH IN THE NEAREST COMPARABLE PERIOD TO 1916-1926 OF VARIOUS
ELEMENTS IN THE SOCIETAL STRUCTURE

Year	Number of nat- ional banks	Savings bank depositors (number)	Year	Consumption of wines and liquors (gallons)	Year	Packs of playing cards on which internal revenue was paid (number)	Cigarettes manufactured weighing not more than 3 pounds per 1000 (number)	Year	Motion- picture manufac- turers (number)	Year	Automobiles registered (number)	Year	Value of radio appa- ratus manu- factured (\$000)
1926	7978	10,950,062	1917	2,095,535,005	1926	42,104,330	92,096,973,926	1925	132	1926	22,001,393	1925	176,990
1916	7579	7,916,990	1916	2,005,835,871	1918	23,355,086	46,656,903,224	1919	50	1916	3,512,996	1921	10,648
Absol- ute													
gain	399	3,033,072		89,699,134		18,749,244	45,440,070,702		82		18,488,397		166,342
Per- centage													
gain	5.3	38.3		4.5		80.3	97.4		164.0		526.3		1562.2

was not free to express overtly any desire or preference it might have for this particular form of diversion.

Between 1918 and 1926 the manufacture of playing cards for domestic consumption increased from 23 million to 42 million packs, or 80.3%. In the same period cigarettes went from 46 billion to 92 billion, an increase of 97.4%.

Between 1919 and 1925 the number of motion-picture manufacturers increased 164%. From 1916 to 1926 the number of passenger automobiles in use increased 526.3%. Between 1921 and 1925, four years, the value of the radio apparatus manufactured in this country increased more than 1500%.

If it be urged that automobiles, moving pictures, and radio apparatus are relatively recent inventions, and therefore would be expected to show rapid percentage growth at this time, the point might be made that their growth is a direct and precise consequence of the psychological fact that the people *want* these products. They could not have shown the growth they did unless people bought them. They were not given away, nor was there any law compelling people to buy them. Furthermore, within the dates covered, there was surely no lack of opportunity, due to the recency of the inventions, for people who desired to go to the motion-picture theaters, or buy automobiles, or radio outfits. And it is difficult to suppose that any such consideration can have had anything to do with the consumption of alcoholic beverages, playing cards, or cigarettes. These mechanisms for diversion were all invented a sufficiently long time ago.

Altogether, it appears that certainly several non-religious forms of diversion, and some non-religious enterprises making for societal stability and conservatism as well, grew in magnitude, and presumably in influence during the decade under consideration much faster than did the religious ones.

IV

It has always been a basic principle in the technique of organized religion that it is desirable to operate vigorously upon the young. A regular and rather elaborate machinery to this end, usually called the Sunday School, is an important element of most church organizations. Among the material gathered by the Census Bureau are included data relative to Sunday Schools. These data are assembled in Table 3.

Table 3 shows that during the decade 1916 to 1926, the number of

TABLE 3

THE NUMBER OF (a) OFFICERS AND TEACHERS, AND (b) SCHOLARS, IN SUNDAY SCHOOLS IN THE UNITED STATES IN 1916 AND 1926

State	Officers and teachers		Scholars	
	1926	1916	1926	1916
United States	2,167,848	1,952,631	21,038,526	19,935,890
<i>New England</i>	<i>100,621</i>	<i>111,073</i>	<i>974,159</i>	<i>1,105,119</i>
Maine	11,851	14,881	106,737	130,433
New Hampshire	6,830	8,272	62,219	77,593
Vermont	5,888	7,487	45,872	61,109
Massachusetts	50,058	53,047	496,375	555,236
Rhode Island	7,425	8,513	85,896	90,393
Connecticut	18,569	18,873	177,060	190,355
<i>Middle Atlantic</i>	<i>360,938</i>	<i>360,633</i>	<i>3,733,158</i>	<i>4,052,130</i>
New York	115,273	118,473	1,151,156	1,296,956
New Jersey	48,980	47,002	489,651	501,756
Pennsylvania	196,685	195,158	2,092,351	2,253,418
<i>East North Central</i>	<i>409,753</i>	<i>386,048</i>	<i>4,163,569</i>	<i>4,028,299</i>
Ohio	125,415	123,153	1,330,818	1,330,312
Indiana	78,738	73,633	773,970	735,967
Illinois	113,246	103,857	1,085,579	1,083,048
Michigan	59,548	54,128	662,687	555,007
Wisconsin	32,806	31,277	310,515	323,965
<i>West North Central</i>	<i>272,732</i>	<i>256,484</i>	<i>2,520,926</i>	<i>2,449,855</i>
Minnesota	37,613	32,723	353,294	328,883
Iowa	55,819	56,005	534,220	540,174
Missouri	71,198	67,893	649,540	664,567
North Dakota	11,689	11,291	103,589	102,411
South Dakota	12,899	10,560	123,270	98,825
Nebraska	28,960	27,879	268,423	253,765
Kansas	54,554	50,133	488,590	461,230
<i>South Atlantic</i>	<i>395,932</i>	<i>337,852</i>	<i>3,884,218</i>	<i>3,408,250</i>
Delaware	5,848	5,825	53,883	56,280
Maryland	34,365	34,193	329,644	329,626
Dist. Columbia	7,717	6,763	83,422	75,094
Virginia	70,060	60,083	688,060	598,379
West Virginia	39,830	34,624	387,312	337,682
North Carolina	91,369	69,599	953,153	739,215
South Carolina	50,538	42,351	507,235	452,047
Georgia	61,742	60,229	567,449	613,947
Florida	34,463	23,685	314,060	205,980
<i>East South Central</i>	<i>231,657</i>	<i>191,964</i>	<i>2,089,136</i>	<i>1,913,790</i>
Kentucky	50,330	44,800	496,109	476,566
Tennessee	60,986	51,131	602,776	523,681
Alabama	67,532	53,612	604,256	524,935
Mississippi	52,809	42,421	385,995	388,608
<i>West South Central</i>	<i>235,005</i>	<i>182,886</i>	<i>2,149,795</i>	<i>1,795,993</i>
Arkansas	45,327	38,238	385,369	342,947

TABLE 3 (continued)

State	Officers and teachers		Scholars	
	1926	1916	1926	1916
Louisiana	28,480	21,698	242,837	221,326
Oklahoma	43,541	31,888	419,295	313,346
Texas	117,657	91,062	1,102,294	918,374
<i>Mountain</i>	<i>60,681</i>	<i>52,734</i>	<i>553,564</i>	<i>494,359</i>
Montana	6,916	6,579	62,587	59,788
Idaho	10,610	9,945	85,580	80,623
Wyoming	3,274	2,716	30,749	23,129
Colorado	17,345	14,181	163,692	139,406
New Mexico	4,153	3,855	40,436	45,122
Arizona	4,571	2,945	44,968	30,884
Utah	12,884	11,731	117,442	108,257
Nevada	928	782	8,110	7,150
<i>Pacific</i>	<i>100,529</i>	<i>72,957</i>	<i>970,001</i>	<i>688,095</i>
Washington	23,264	20,866	219,216	184,630
Oregon	15,301	13,786	142,997	120,162
California	61,964	38,305	607,788	383,303

Sunday-School scholars increased only 1,102,636 in absolute numbers, or 5.5%, as against an increase in population of 16.3% in the same period. At the same time the number of officers and teachers in Sunday Schools increased 215,217, or 11.0%. There was thus evidently no lack of effort to bring the young into the fold, but the effort was plainly not marked by the most signal success.

Taking the separate geographical divisions, it is to be noted first that in New England—the home of our Puritan traditions—the number of Sunday-School scholars actually *decreased* during the decade by 130,960 souls or 11.9%. In that section of the country apparently even the effort is weakening, because the number of officers and teachers in Sunday Schools decreased by 10,452 head, or 9.4%.

In the Middle Atlantic states, again, the number of Sunday-School scholars *decreased* by 318,972 souls, or 7.9% of the number enrolled in 1916, in spite of the fact that there were 305 more officers and teachers in 1926 than in 1916.

In the East North Central states the number of Sunday-School scholars increased by 135,270, or 3.4%, and in the West North Central states by 71,071, or 2.9%, both small and obviously insignificant increases.

On the other hand, the number of Sunday-School scholars increased in the South Atlantic states by 475,968, or 14.0%, in the East South

Central states by 175,346, or 9.2%, in the West South Central states by 353,802, or 19.7%, in the Mountain states by 59,205, or 12.0%, and in the Pacific states by 281,906, or 41.0%.

Now the New England, Middle Atlantic, East North Central, and West North Central geographical divisions of our country include 21 states, which had in 1920, taken together, a population of 63,681,845, or 60.2% of the total population of the whole country. On the other hand, the remaining divisions (South Atlantic, East South Central, West South Central, Mountain, and Pacific) include 27 states and the District of Columbia and had, taken together, a population in 1920 of 42,028,775 or 39.8% of the total population of the whole country. So then it appears that, in those parts of the country which include 60.2% of the whole population, Sunday Schools made no significant growth whatever in a decade during which the country as a whole grew 16.3% in population. But the rest of the country, with only 39.8% of the total population, gathered into the fold during the same period, or at least had, at the end of the period in 1926, 1,346,227 more Sunday-School scholars than they had in 1916.

It is further plainly apparent that it is in the older and more highly developed sections of the country where the number of Sunday-School scholars either has increased by insignificant percentages in the decade from 1916 to 1926 or has actually decreased. In the newer and less developed sections of the country, on the other hand, there have been substantial increases. Apparently, this is another instance of the trend noted by Whitehead. It probably may be expected that as the civilizations of South and West evolve farther, the influence of their Sunday Schools will begin to decay.

In the New England, Middle Atlantic, East North Central, and West North Central states, the number of illiterate persons 10 years of age and over in 1920 was 1,843,773, or 3.6% of the total population in this age class. For the South Atlantic, East South Central, West South Central, Mountain, and Pacific states the corresponding figures were 3,088,132 and 9.6%. In other words, it is in general in the more illiterate sections of the country that the number of Sunday-School scholars shows a substantial increase, whereas in the better educated sections it has increased by insignificant amounts or has decreased.

V

Up to this point the discussion has had to do with all religious denominations lumped together. In 1926 there were 213 separate

denominations reported. It is impossible in this brief discussion to consider all of them separately, but in Table 4 twenty-five have been picked out which are of interest for one reason or another, and these are arranged in descending order of the percentage gain in membership during the decade 1916-1926. A plus sign (+) indicates a gain and a minus sign (—) a loss. Some of these denominations in Table 4 are small and statistically unimportant ones, but these 25 denominations had a total membership of 34,502,495 in 1926, or 63.2% of the total church membership reported in the 1926 census. Thus these 25, which constitute only a little more than 10% of the denominations, include more than 60% of all church members.

TABLE 4
PERCENTAGE GAIN OR LOSS IN MEMBERSHIP IN 25 DENOMINATIONS
BETWEEN 1916 AND 1926

Name of denomination	Percentage gain (+) or loss (—) in period 1916 to 1926
Church of God	+198.7
Church of the Living God, "The Pillar and Ground of Truth"	+190.9
Pilgrim Holiness Church	+185.1
Volunteers of America	+181.8
Christian and Missionary Alliance	+136.2
Polish National Catholic Church	+118.0
Pillars of Fire	+116.2
National Spiritualist Association	+ 77.8
General Conference of the Mennonite Church of North America	+ 40.1
Seventh Day Adventist Denomination	+ 39.9
Churches of Christ	+ 36.4
Southern Baptist Convention and American Baptist Association	+ 34.5
Old Catholic Churches in America	+ 27.1
Protestant Episcopal Church	+ 25.5
Roman Catholic Church	+ 18.3
United Lutheran Church in America	+ 17.9
Methodist Episcopal Church, South	+ 17.7
Methodist Episcopal Church	+ 9.8
Northern Baptist Convention	+ 8.3
Society of Friends (Orthodox)	— 1.1
Universalists	— 6.2
Evangelical Synod of North America	— 7.5
Colored Methodist Episcopal Church	— 17.5
Christian Union	— 35.8
Two-Seed-in-the-Spirit Predestinarian Baptists	— 55.2

The "Church of God," which heads Table 4, as showing the greatest relative growth in the decade 1916-26, flourishes in the South. Its intellectual level is indicated by the following statement of doctrine (6, Vol. 2, p. 361).

"In doctrine this body is Arminian and in accord with the Methodist bodies. It recognizes no creed as authoritative, but relies upon the Bible 'as a whole rightly divided' and as the final court of appeals. It emphasizes sanctification as a second definite experience subsequent to regeneration; also the baptism of the Holy Ghost, evidenced by speaking in other tongues, subsequent to sanctification. Conditions of membership are profession of faith in Christ, experience of being 'born again,' bearing the fruits of a Christian life, and recognition of the obligation to accept and practice all the teachings of the church. The sacraments observed are the Lord's Supper, foot washing, and water baptism by immersion."

The second on the list of Table 4, with a percentage gain of 190.9% in the decade, is the Church of the Living God, "The Pillar and Ground of Truth," a body which split off in 1902 from a more ancient ecclesiastical organization founded in Arkansas in 1889. Its distinctive doctrinal characteristics are baptism by immersion, washing of the saints' feet, and the use of water and unleavened bread in the celebration of the Lord's Supper.

The third in the list, the Pilgrim Holiness Church, made a gain of just over 185% in membership during the ten years and had in 1926 a total membership of 15,040. Each applicant for membership in that church is "questioned in the presence of the church at a public meeting," before admission, as to his or her complete acquiescence in the articles of the faith of the church, which emphasize, among other things, belief that the Holy Scriptures are "divinely and supernaturally inspired, infallibly true as originally given, and as the only divinely authorized rule of faith and practice." While it is not expressly so stated, "originally given" probably refers to the King James English edition of the Bible.

The Volunteers of America, with a gain of 181.8 per cent, is a church auxiliary organization including adherents of various denominations. The Christian and Missionary Alliance, with a gain of 136.2 per cent, is strictly evangelical in doctrine and stands firmly for the inspiration of the Scriptures.

Considering now the larger denominations, it is interesting to note that the Southern Baptists made the largest relative gain in member-

ship, with the Protestant Episcopal Church at the next place in the table, when adjustment is made for changes in definition of membership between 1916 and 1926. The Methodists (of the South), the Lutherans, and the Catholics grew at about the same rate. The Methodists (of the North) and the Northern Baptists, however, made but relatively small gains.

The Quakers lost slightly. While but one particular denomination of the sort is shown in Table 4, the census reveals the fact that the Negro Methodist denominations tended generally to lose in membership in the decade. The reason for this appears fairly obvious. The Christian Union, a doctrinally liberal body, lost rather heavily in membership. It is impossible to refrain from a word of condolence over the great loss in membership of so sturdy a body as the Two-Seed-in-the-Spirit Predestinarian Baptists. Any organization with a name like that deserves better of its fellow men.

VI

Such a cursory examination as this is of a large mass of data cannot possibly lead to wide or definitive generalizations. Indeed, the chief purpose is only to call the attention of students of social psychology to the wealth of interesting material which is available in the Census of Religious Bodies, 1926.

So far as concerns the particular matters here discussed, one gets distinctly the impression from a first examination of the available data that there is a tendency for those religious denominations and sects which are characterized by a narrow and rigid body of doctrine, and therefore appeal to the less intelligent sections of the population, to grow more rapidly than those whose more liberal doctrines appeal to the more intelligent sections, if at all. Not a few of the latter denominations seem to be losing in membership and presumably, therefore, influence, at a rapid rate. It must always be remembered, however, that *more than half of the population of the United States* in 1926 had no church affiliations whatever. This is a great basic fact. It should be encouraging to both the religious and the infidels. For the former it is a stimulus. There is work yet to be done. For the latter it furnishes a pleasant morsel for contemplative rumination.

The inverse relationship suggested by the data between the vigor of the churches and the general intellectual and cultural level of the population is perhaps to be expected. Benda (2) points out in a recent volume that all of our political and philosophical ideas, which

we fondly believe to be the outcome of our reason and our judgment, are in reality only adopted because they satisfy our "*dispositions sentimentales*." Probably he is right. The "*dispositions sentimentales*" of old civilizations are different in many respects from those of newer and cruder ones.

In this same connection, it may be noted that, in the course of a recent study of the incomes and disbursements of 96 members of the University of California faculty, Peixotto (5) makes the following statement:

"It is interesting to report, though difficult to explain, the fact that the fifty-two members of the group who recorded church contributions were those with small expenditures and that it was generally true that the percentage supporting churches decreases as total expenditure rises."

REFERENCES

1. [ANON.] One year in the ministry. *Scribner's*, 1929, **86**, 629-639.
2. BENDA, J. Mon premier testament. *Paris* (nrf.) 1928.
3. MCCORMICK, W. P. G. Be of good cheer. London: Longmans. 1930.
4. PEARL, R. The psychology of prohibition. *Amer. Mercury*, 1928, **15**, 385-394.
5. PEIXOTTO, J. B. Family budgets of university faculty members. *Science*, 1928, **68**, 497-501.
6. UNITED STATES DEPARTMENT OF COMMERCE. Bureau of the Census. Religious Bodies, 1926, (2 vols.) Washington, U. S. Govt. Printing Office; 1930.
7. WHITEHEAD, A. N. Science and the modern world. New York: Macmillan, 1925. Pp. xi+296.

The Johns Hopkins University
Baltimore, Maryland

QUELQUES NOTES SUR LE RECENSEMENT DES GROUPES RELIGIEUX, 1926

(Résumé)

Cet article discute les résultats du recensement des groupes religieux aux Etats-Unis en 1926 fait par le Bureau de Recensement. On montre que tandis que la proportion de membres de l'église est devenu un peu plus grande en croissance que celle de la population entière en les dix ans 1916-1926, cette proportion en ces dix ans a été plus petite que celle des clients des caisses d'épargne, celle du nombre de cartes à jouer pour lesquelles on a payé la contribution indirecte, celle des cigarettes employées, celle des automobiles enregistrées, celle du nombre de fabricants de films, et celle des postes de radio fabriqués. On conclut que plusieurs formes non religieuses

de diversion et des entreprises non religieuses tendant à causer la stabilité sociale sont devenues plus grandes et de plus d'influence, on peut le supposer, en les dix ans considérés, plus vite que les formes religieuses.

Dans les régions plus âgées et plus lettrées des Etats-Unis, le nombre des élèves des écoles religieuses de dimanche est devenu seulement un peu plus grand dans les dix ans considérés, et dans quelques grandes sections du pays le nombre de ces élèves est devenu plus petit, malgré la croissance de la population. Dans les régions nouvelles et moins lettrées le nombre en est devenu beaucoup plus grand.

Il paraît que les dénominations et sectes religieuses qui possèdent comme trait caractéristique un dogme étroit et rigide et donc attirent les parties les moins intelligentes de la population tendent à croître plus rapidement que celles dont les doctrines plus libérales attirent les parties les plus intelligentes.

PEARL

EINIGE BEMERKUNGEN ÜBER DIE SCHÄTZUNG RELIGIÖSER GRUPPEN, 1926

(Referat)

Diese Abhandlung bespricht die Befunde bei der Abschätzung der religiösen Gruppen in den Vereinigten Staaten die im Jahre 1926 von dem Zensusbureau gemacht wurde. Es zeigt sich, dass obwohl die Kirchenmitgliedschaft im Allgemeinen sich während des Jahrzehntes 1916 bis 1926 etwas mehr vermehrte als die Bevölkerung im Ganzen, die Geschwindigkeit des Wachstums (rate of growth) dieser Mitgliedschaft während dieser Periode doch geringer war als die bei der Zahl der Einzähler in Sparkassen, bei der Zahl der Pakete von Spielkarten die von dem Staate taxiert wurden, und bei der Zahl der verbrauchten Zigaretten, der registrierten Automobile, der Filmfabrikanten, und der fabrizierten Radioapparate. Der Verfasser schliesst hieraus, dass verschiedene nicht-religiöse Unternehmungen die die soziale Festigkeit begünstigen während der berücksichtigten Periode in ihrer Grösse und wahrscheinlich auch in ihrem Einfluss viel rascher zugenommen haben als die religiösen.

In den älteren und mehr gebildeten Teilen der Vereinigten Staaten nahm die Zahl der Sonntagsschüler (d.h., der religiöse Unterricht) während des berücksichtigten Jahrzehntes nur unbedeutend zu, und in einigen gossen Teilen des Landes fand sogar eine Abnahme in der Zahl der Sonntagsschüler statt, trotz der Bevölkerungszunahme. In den jüngeren und minder kultivierten Teilen war die Zunahme beträchtlich.

Die Daten scheinen darauf hinzuweisen, dass diejenigen Konfessionen und Sekten die sich durch eine beschränkte und strenge Lehre auszeichnen und deshalb besonders dem minder intelligenten Teil der Bevölkerung zusagen, etwas schneller wachsen, als diejenigen, deren freisinnigere Lehren eher den intelligenteren Teilen zusagen.

PEARL

PSYCHOLOGICAL STUDIES OF MOTION PICTURES:
V. ADOLESCENT AND ADULT SEX DIFFERENCES
IN IMMEDIATE AND DELAYED RECALL*

From the Institute of Child Welfare of the University of California

HERBERT S. CONRAD AND HAROLD E. JONES

INTRODUCTION

Prior studies in this series (5, 10) have illustrated the use of motion pictures in the study of individual differences in learning and recall. Tests of observation and memory for events have usually been restricted to "staged" activities, as in the conventional experiment on testimony; the screened episode, however, offers certain advantages over the older technique. The film makes possible an accurate control of tempo, an almost unlimited range of event sequences, and a precise repetition or an experimental variation of either conditions or content. It also presents advantages in the motivation of subjects and in the appeal to relatively unselected samples (6). Boring (4) appears to have been the first to adopt the use of film material in the study of testimony; on the basis of a one-minute picture he obtained a narrative and deposition from those observing the picture, and concluded that "there is considerable presumption of a difference in excellence of report between men and women in favor of the former; that no such sex difference is apparent in childhood; and that the reports of adults are more adequate and accurate than those of children." These and other conclusions were derived from a sample consisting of 11 men, 13 women, 12 boys, and 8 girls.

In the course of an investigation¹ conducted by the writers in rural New England, reports on the content of three motion pictures were obtained from 746 subjects, including 353 adolescents (10-17

*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office, November 13, 1930.

¹The survey, of which this is a part, was conducted under a grant from the Social Science Research Council of Columbia University. Acknowledgments are due to Professors R. S. Woodworth and A. T. Poffenberger, of Columbia University, for advice and administrative assistance; and to the Yale University Press for the loan of a film from the "Chronicles of America" series.

years) and 393 adults (18-54 years). Of these records, 92 were obtained after a delay period of one week; the remainder were the result of an immediate test following the presentation of the picture. The films employed, and the method of preparing the tests, have been described in a prior publication (10, pp. 226-227).² The tests were given in eight villages in Vermont, selected, by a number of criteria, as representative of rural communities in that state. The average size of the audiences participating in the tests was about 75; copying was prevented by the use of alternate forms (the same questions in a slightly changed order) for persons sitting in adjacent seats. It is probable, however, that this was an unnecessary precaution, as each community had been brought to an excellent degree of rapport with the purposes of our survey. Methods of administration, and of testing the representativeness of the sample, have been elsewhere reported in some detail (5, 6, 10). A small number of tests were eliminated in making up our present group of 744 cases. These rejected records were from persons of foreign birth and from transients and summer visitors; the former were eliminated because of a possible language handicap in dealing with verbal tests, the latter because they represented an unverifiable sampling.

²The films were:

A. A picture from the Yale Historical Series, describing the defeat of the British in the West by Clark's army of frontiersmen. (*Vincennes*, modified to two reels.)

B. A romantic drama, dealing with "character" narrative in a setting of (1) a small country town, and (2) professional league baseball (*The Busher*, modified to four reels). Three popular stars of the period were featured (Charles Ray, John Gilbert, Colleen Moore).

C. A "western" picture (featuring Hoot Gibson), consisting chiefly of the prevailing cowboy rituals of "action," and also containing a minor love story (*The Loaded Door*, five reels).

"On the subject-matter of each picture, tests were made consisting of from 40 to 48 completion and multiple choice items; the questions dealt with verbal as well as with pictorial material, and with incidental episodes and atmosphere as well as with direct continuity. Each reel was systematically covered, the final test being edited by a group of psychologists who had studied the film in experimental previews.

"It was desired that individual differences in test scores should depend basically upon differences in the observation, understanding, and recall of the picture. This aim was sought by making the questions of a nearly zero range of difficulty in vocabulary and phrasing. Each question was brief, direct, and simply worded. Within the educational limits of our group, variable test performance may be considered as chiefly a function of 'picture comprehension and memory,' and not of 'paper and pencil ability.' (Of the adults taking the tests 98 per cent had completed grade 8 of elementary school)."

In comparing our procedure with Boring's, certain differences are evident. Since we were testing not on a one-minute presentation but on screen material requiring from 30 to 75 minutes to exhibit, it was scarcely possible to employ a narrative form of report, nor was it feasible to obtain statements on each item with regard to "assurance" or "willingness to attest." In the conventional terms of coefficients of report of a testimony experiment (21), our test scores may be regarded as $n(D)$, absolute range of report in the deposition. The relative range of report may be obtained from $n(D)$

$\frac{n(D)}{p}$, where p represents the total number of items in the test.

In the terms commonly employed in mental testing, relative range of report equals the subject's actual score, divided by the total possible score.

In group measurements of adults, a time limit is ordinarily necessary (6, p. 281); our time limits were set at 11 minutes for Test A, 14 minutes for Test B, and 7 minutes for the part of Test C which is appropriate for the present report. With this timing, about 5% finished Test B, 20-25% finished Test A, and 80-85% finished Test C. Because of time restrictions, the subject's score is, of course, to some extent dependent upon speed, and in the latter part of the test no clear distinction can be made between questions unanswered through lack of time or through lack of knowledge. This may be considered an irrelevant or disturbing factor in the present study, which is concerned with the power of recall rather than with the speed of registering testimony. The speed factor, however, has been controlled, where necessary, by comparisons in terms of position in series (pages 437-438).

In Boring's study, the subjects were kept in ignorance of the fact that an interrogatory would follow the picture. In the present investigation, the spectators knew that some sort of a test was to be given during the evening, but they were not told that this was to be based on the content of the motion picture. There is no doubt that our typical rural audiences were interested in the motion picture simply for the entertainment afforded, rather than for the sake of any test which might follow. Only one motion-picture test was given within a single performance. In some communities, however, a return was made a week or so later with a second picture. This has resulted in a slight overlapping between the groups taking Tests A,

B, and C. Thus, of the 294 individuals tested on motion picture A, 93 were also subsequently included in those taking the test on motion picture B.

SEX DIFFERENCES

The motion picture tests were marked by one of the authors according to a written standard, prepared from an extensive sampling of replies. A correct answer on a test question counted 2, an omission or wrong answer counted 0; half-credits of 1 were also occasionally assigned, according to standard.

Throughout the present study, the individuals aged 10-17 will be considered as a single group—viz., “adolescent”; those aged 18-54 will, similarly, be considered “adult.” Sex differences will always be considered within these groups.

In Table 1 no significant sex differences in variability appear. In means and medians, however, a slight difference is apparent among

TABLE 1
MEANS, MEDIANS, AND STANDARD DEVIATIONS OF TOTAL SCORES*

Sample	Adolescent Groups (Age 10-17)				Adult Groups (Age 18-54)			
	<i>n</i>	Meant	Median†	<i>S.D.</i> _{cts.}	<i>n</i>	Meant	Median†	<i>S.D.</i> _{cts.}
Test A:								
Male	69	43.2±1.57	46.6	19.3	75	51.5±1.43	52.0	18.4
Female	70	42.8±1.36	42.0	16.9	80	46.0±1.42	46.0	18.9
Test B:								
Male	49	35.3±1.82	34.9	18.9	43	51.5±1.93	51.3	18.8
Female	57	42.1±1.66	39.2	18.6	56	48.3±1.75	47.5	19.4
Test C: _{immed.}								
Male	30	29.4±1.02	31.5	8.3	46	33.2±.73	36.0	7.3
Female	32	28.7±1.08	30.7	9.0	47	31.5±.83	33.9	8.5
Test C: _{delay}								
Male	22	22.1±1.63	22.5	11.3	20	29.4±1.18	33.2	7.8
Female	24	23.8±1.62	25.0	11.8	26	28.1±1.03	29.8	7.8

*The reliability of the total test scores is from .90—.94 in groups closely comparable to those in the present study (5, pp. 260-261).

†The figures in this column represent central tendencies of the raw scores (“absolute range of report”). The “relative range of report” may be obtained by dividing the figures given by the maximum score possible for the test in question; namely, for Test A, 80; for Test B, 96; for Test C, 40. Note that in Tables 2, 5, 6, and 7, the figures represent relative range of report.

the adults, the males being from one-sixth to one-third *S.D.* superior to the females. This adult male superiority, while in no particular instance statistically reliable, is nevertheless sufficiently consistent to raise seriously the question of a genuine adult male superiority in recall or capacity to report. Three problems need to be considered in this connection: (*a*) the possible existence of a "speed" difference between males and females, which would enter into recall scores by handicapping one sex more than the other in the completion of the tests; (*b*) the possibility of a sex difference in general intelligence; and (*c*) the possibility that our pictures appealed more effectively to the interests of one sex than to those of the other.

The Speed Factor

The influence of a speed factor may be examined by comparing scores on the first third and the last third of each test. The first third was completed by all, but because of the time limits many failed to reach the last third of Test B or of Test A. (As previously stated, 80-85% finished Test C.)

Table 2 shows that, among adults, the superiority of males, in Tests A and B, is substantially greater in the first than in the last third of the test. Since these sections are closely comparable in content and in the types of items represented, the difference cannot be accounted for in terms of a greater masculine appeal of the first part of each picture; apparently the difference is due chiefly to the fact that the adult males work more slowly. While the beginning of the test measures primarily the power of recall (dependent, of course, on the amount learned), toward the end of the test a speed factor becomes increasingly apparent, with a greater handicap to adult males

TABLE 2
THE INFLUENCE OF THE SPEED FACTOR

Test	Adolescents				Adults			
	Percentage of correct answers in				Percentage of correct answers in			
	First third of the test	Last third of the test	First third of the test	Last third of the test	First third of the test	Last third of the test	First third of the test	Last third of the test
	Male	Female	Male	Female	Male	Female	Male	Female
A	62	62	44	42	72	61	54	53
B	52	57	22	26	66	57	33	42
<i>C_{immed.}</i>	78	71	75	79	85	84	85	82
<i>C_{delay}</i>	60	62	59	51	83	75	74	70

than to females.³ Correction for this handicap would, of course, increase the sex differences among the adults in Table 2. Up to this point, then, Boring's finding of an adult male superiority has been confirmed. But consideration of the other two factors (intelligence and interest) will raise strong doubts of the validity of this difference.

The Intelligence Factor

At one time or another during the survey, approximately one-half of our motion-picture sample was tested by a group intelligence test (Army Alpha, Forms 5 or 7). In a previous article the writers have reported correlations of from .6 to .7 between the average motion-picture test and the Army Alpha.⁴ This correspondence is sufficiently close to lead us to predict differences in motion-picture scores in groups showing any marked inequivalence in Alpha. Table 3 presents the Alpha-tested subsample of the cases reported in Table 1; it is clear from the correspondence (in Table 3) between the median age of the subsample and the principal sample, and the median motion-picture scores of the two samples, that the subsample may be taken as fairly representative of our principal group. In the subsample, it is noteworthy that the outstanding sex differences in motion-picture scores are accompanied by outstanding differences in Alpha. Thus, among the adolescents in Test *C_{delay}*, the female superiority of about .9 *S.D.* in motion-picture score is paralleled by a superiority of about .8 *S.D.* in Alpha.⁵ Among the adults in Test A, the male superiority of 1 *S.D.* in Test A parallels a superiority of about .7 *S.D.* in Alpha; and the male superiority of .5 *S.D.* in Test *C_{delay}* parallels a superior-

³In Test C (especially Test *C_{immediate}*), the speed factor appears to be of little importance for either sex; here, as previously stated, the questions were markedly easier in relation to the time allowance than in the case of the other two tests.

From evidence external to the present study, it seems probable that if our sample were restricted to rural adults above the age of 40, the effect of the speed factor in Tests A and B would be intensified.

⁴For adults, Test A x Alpha, .67; Test B x Alpha, .69. For adolescents, with age partialled out, the comparable figures are .71 and .62 (5, p. 254). On the basis of a study of normal children, McGeoch has stated that "the correlations between intelligence and report-ability... show a tendency to zero or inverse relation..." (11, p. 444). This absence of positive relation is doubtless due largely to the unreliability of McGeoch's measure of an individual's "report-ability," since his results are based on uncombined scores for extremely short exposure periods (20-30 seconds).

⁵The standard deviation in Army Alpha scores for groups of adolescents is about 33, for groups of adults, about 43 (5, pp. 252-253).

TABLE 3
THE INFLUENCE OF THE INTELLIGENCE FACTOR

Test	Sex	No. of cases		Median age		Median motion-picture test score		Alpha score of subsample
		In principal sample	In sub-sample	Of principal sample	Of sub-sample	Of principal sample	Of sub-sample	
Adolescent Groups								
A	Male	69	32	14.0	13.7	46.6	43.0	78.0
	Female	70	29	14.3	14.1	42.0	43.5	86.5
B	Male	49	32	13.9	13.8	34.9	36.5	82.5
	Female	57	27	14.3	14.3	39.2	40.5	85.5
C _{immed.}	Male	30	13	14.5	15.8	31.5	33.0	87.0
	Female	32	12	15.5	15.0	30.7	34.0	75.0
C _{delay}	Male	22	14	13.5	13.5	22.5	21.0	59.0
	Female	24	9	13.8	13.6	25.0	31.0	85.0
Adult Groups								
A	Male	75	22	26.5	26.3	52.0	60.5	114.0
	Female	80	24	34.2	37.7	46.0	42.0	85.5
B	Male	43	19	31.1	34.0	51.3	48.5	117.5
	Female	56	33	34.3	31.0	47.5	48.5	108.5
C _{immed.}	Male	46	22	24.7	31.0	36.0	35.5	106.0
	Female	47	19	34.4	34.0	33.9	36.0	96.0
C _{delay}	Male	20	11	30.8	34.0	33.2	33.0	119.0
	Female	26	7	40.8	38.0	29.8	29.0	97.0

ity of .5 *S.D.* in Army Alpha. The consistent superiority of the adult males over the adult females in Alpha scores is not a reflection of a general sex difference in the rural communities selected for the present study; on the contrary, it appears to be characteristic only of this particular sampling, and is probably due to the fact that our motion-picture attendance included a slightly more representative selection of adult females than of adult males.

It is evident, then, from the subsample, that the handicap of our adult males in the speed of test performance is counteracted by an advantage in intelligence.⁶ One may wonder to what extent previous

⁶The males of our sample also enjoy an advantage in chronological age [cf. Table 3; data on the decline of motion-picture test scores with age are reported in references (5, pp. 255-258, and 10, pp. 233)]. The failure of intelligence differences between the sexes, within our sample, to parallel speed differences in motion-picture test performance, suggests an interesting field for further investigation. Within each sex, and in both sexes taken together, speed and Alpha score are, of course, positively correlated.

findings of sex differences (1, 2, 4, 16, 20, 22) are due to similar sampling differences in intelligence arising either through chance or through systematic factors (16). If, within our sample, we were to match pairs on the basis of intelligence, the slight superiority of the adult males would certainly be reduced, or perhaps eliminated. Before, however, attempting this, it will be wise to consider the influence of a third factor, namely, the possibly differential appeal of our motion pictures for the two sexes.

The Interest Factor

A number of studies (3, 13, 15, 17, 19) give evidence in support of the contemporary belief that we remember most readily what is most interesting to us or is most closely related to our past experience. This tendency may in part be dependent upon emotional orientations, which vary widely among individuals, and undoubtedly involve certain characteristic sex differences. Smith (18) and Jones (9) have studied the relationship between the emotional reactions to verbal stimuli, as registered by a galvanometer, and their recall values, as indicated in a memory experiment; both of these investigators have been able to show that positively toned words are remembered more easily than words of small emotional significance, and that negatively toned words are, on the average, poorly remembered. It is not unlikely that certain stimuli may, on the average, have an opposing emotional polarity for males and females, with corresponding opposed effects upon learning and memory.

From these considerations, a study of sex differences in the recall of *individual items* would seem to offer more promise than an investigation restricted to total range of report, for it is possible that sex differences in total score (even when all other ability factors are controlled) reflect no differences in such a supposed generic trait as "power to recall," but depend merely upon the averaging of many specific responses, and ultimately upon the particular assortment of items which have been brought together. An obvious corollary of this is that true sex differences may occur for individual items, but may vanish, or dwindle into insignificance, when added algebraically (in) the total test score.

Table 4 gives, for each picture, the number of test items for which we found sex differences of a fairly reliable nature (the chances are 90 in 100 or greater, in favor of a true difference).⁷ In Table 4,

⁷The table was obtained as follows: On item number 2 of Test A, 74%

TABLE 4
SEX DIFFERENCES IN INDIVIDUAL TEST ITEMS*

Test (1)	No. of reels in picture (2)	No. of items in test (3)	Percentage of items showing a sex difference		Adolescent group		Adult group	
			Adoles- cents (4)	Adults (5)	Male superi- ority (6)	Female superi- ority (7)	Male superi- ority (8)	Female superi- ority (9)
A	2	40	15	43	5	1	17	0
B	4	48	38	52	2	16	13	12
<i>C_{immed.}</i>	5	20	20	45	3	1	8	1
Total	11	108	28	47	10	18	38	13

*Test *C_{delay}*, which duplicates the content of Test *C_{immed.}*, and was taken by fewer cases, has been omitted from this table. The set of results for *C_{delay}* is similar to that for *C_{immed.}*, except for a smaller male superiority.

the extent of sex differences in response to individual test items may be misunderstood unless the following points are noted: (a) a certain number of differences between the adult males and females is to be expected as an outcome of the sex differences in speed and in intelligence already noted; (b) by the standard adopted (only 90 chances in 100, not complete certainty), we should expect 10% of the 90/100 differences to be merely chance, and not at all true differences (7, p. 93); this 10% would consist half of questions showing male superiority, and half of questions showing female superiority. It can be seen in Table 4, however, that among the adults the number of questions fulfilling our statistical standard is not 10%, but 52, 43, and 45%; and the division of superiority between the adult males and females is not equal, but markedly unequal. It is safe to infer, therefore, that sex differences in response to individual test questions do exist among adults.

In the case of the adolescents, the evidence for sex differences consists chiefly in the distinctly uneven division of superiority between males and females (cf. Table 4, Columns 6 and 7). Test B is the chief contributor to adolescent sex differences in favor of the females. Items showing a male superiority are much more numerous among

of the adult males answered correctly, 64% of the adult females answered correctly. The difference is 10%, with a *P.E.* of .05. The chances of such a difference being a true difference are over 90 in 100 (7); and this question accordingly is included among those showing a fairly reliable sex difference.

adults than among adolescents, while items showing female superiority are slightly less numerous among adults. The mean of all the adolescent differences (without regard to sign) in Tests A, B, and *C_{immediate}* is 8.8%; of the adult differences, 10.45%.

To gain insight into the type of question in which males or females, respectively, excel, we shall be on a safer ground if we examine only those questions showing a reliability of 98/100 or greater. A verbatim list of such questions, with correct answers indicated, is presented below for each motion picture. In the completion items, our standards, of course, allowed for a variety of alternative, equivalent replies, as well as (on certain items) for replies to which a partial credit could be assigned.

SEX DIFFERENCES ON SPECIFIC ITEMS

	Question from test	Percentage correct Male	Female
ADOLESCENT GROUPS: MALE SUPERIORITY			
1. From Kaskaskia to Vincennes, how far did Clark have to march? 500 miles a day's journey X240 miles nobody had measured it	A	87	75
2. Ben practiced pitching in his back yard. What happened when he threw the ball? (<i>The target broke.</i>)	B	30	14
3. Describe the new clothes he bought. (<i>Straw hat, gray suit, and cane.</i>)	B	79	25
4. Why did Bert leave the ranch? (<i>To make his fortune in the city.</i>)	C _{immed.}	67	39
5. How many months did he stay away? (<i>Ten months.</i>)	C _{delay}	77	45
ADOLESCENT GROUPS: FEMALE SUPERIORITY			
6. Ben had saved \$3.50 for (<i>a new baseball glove</i>).	B	79	94
7. At the shadow social, the girls were not all (<i>"what they seemed"</i>).	B	13	36
8. An auctioneer stood in front of the screen holding (<i>a basket</i>).	B	70	87
9. What did Mazie and the milliner wear on their heads? (<i>Large bows.</i>)	B	42	84
10. What nicknames were given Ben? (<i>Chesty and Hick.</i>)	B	28	56
11. Why did Mazie ask her brother to lend her a dollar? (<i>To go to the baseball game.</i>)	B	75	89

	Question from test	Percentage correct	
		Male	Female
12. Brady complained because (<i>Ben was late and delaying the game</i>).	B	29	47
13. When the Deacon offered Mazie something to eat, why didn't she accept it? (<i>She was sad because of Ben's conduct.</i>)	B	8	28
14. How did Ben come home? (<i>By "riding the rods"</i>).	B	40	58
15. What happened immediately after Ben won the game? (<i>His team swung him on their shoulders.</i>)	B	4	16
16. At the end of the game, the postoffice clerk walked home with (<i>the milliner</i>).	B	5	14
ADULT GROUPS: MALE SUPERIORITY			
17. Why did the family leave Virginia? to visit friends ×to pick and choose their acres	to see Niagara Falls to spy on the Indians	A	87
18. Clark advised the father of the family: Take plenty of blankets Don't visit the British	×Wait till the trails are safe Come back soon	A	90
19. How did the Indians greet the British? By shaking hands By falling on the ground	×By raising their hands By giving a war whoop	A	76
20. Clark asked the Governor of Virginia to (<i>help him with money and troops to drive out the British</i>).		A	72
21. When Clark's men heard what the Spanish trader said, they didn't be- lieve him ×they agreed with him	they said, "let's go home" they smoked a pipe of peace	A	74
22. Describe exactly how he showed emotion when he said this. (<i>He thrust forward his chin and pounded the table.</i>)		A	36
23. The first British soldier who was hit ran away returned the shot	×fell to the ground fixed his bayonet	A	55
24. Describe the new clothes he bought. (<i>Straw hat, gray suit, and cane.</i>)		B	40
25. What happened to the baseball? (<i>It rolled down a chute back to the pitcher.</i>)		B	72

	Question from test	Percentage correct	
		Male	Female
26. He left the yard quietly because (<i>the Deacon was fast asleep</i>).	B	53	33
27. The fence gate was fastened with (<i>a chain</i>).	B	72	50
28. At the baseball game, a spectator said, "That guy's got more curves than (<i>a stove-pipe</i>)".	B	75	52
29. After the baseball game, a man who was pleased with the pitching said, "I may be (<i>able to give you a contract</i>)".	B	76	56
30. While Ben was making his last bid, what was his rival doing? (<i>Looking behind the screen.</i>)	B	79	53
31. When he entered the hotel, what did his team-mates do? (<i>"Kidded" him.</i>)	B	87	68
32. Brady complained because (<i>Ben was late and delaying the game</i>).	B	70	53
33. Coming home Bert rode in what kind of a car? (<i>A freight car.</i>)	C _{immed.}	99	86
34. What happened to Bert when he went to take over the ranch? (<i>He was attacked by Blackie's gang.</i>)	C _{immed.}	80	45
35. Molly's brother drew a gun, waiting for a chance to help (<i>Bert</i>).	C _{immed.}	80	62
36. Bert was out of a job. Who met him on the street and loaned him some money? (<i>The sheriff.</i>)	C _{delay}	100	85
ADULT GROUPS: FEMALE SUPERIORITY			
37. How much was Ben's first bid at the shadow social? (<i>50 cents.</i>)	B	43	80
38. What did Mazie and the milliner wear on their heads? (<i>Large bows of ribbon.</i>)	B	77	94
39. When Ben came home penniless, what did the Deacon shove under the door for Ben? (<i>An envelope containing money.</i>)	B	54	72
40. The Deacon said, (<i>"I kept your baseball glove for you"</i>).	B	31	59
41. How was Ben received by the young folks when he first returned home? (<i>They snubbed him.</i>)	B	37	58
42. Ben said, "This time it's going to be tickets for (<i>two—for Mazie and me</i>).	B	28	46

From the above list of sex differences, two facts emerge fairly clearly: (*a*) each sex excels in questions touching their peculiar interests or training; (*b*) each sex is interested in (or at least recalls

better than the other sex) its own characteristic activities or duties. It is evident that the normal "heterosexual fixation" does not imply a more ready recall for the situations and events which are primarily associated with the other sex. In Test A the striking absence of a single case of female superiority (of reliability 98/100), either among the adolescents or adults, we interpret as due at least in part to the absence of any leading feminine characters in this picture, and the lack of a love story or other specific feminine appeal. Test A is based on a picture about the conquest of the Northwest, with masculine exploits providing the sole theme. In Test B, on the other hand, the prominence of a heroine ("Mazie"), and of a love story, appears to raise the level of feminine recall very markedly. Among the adolescents, where the females are perhaps sooner or more keenly aware of and interested in romantic and distinctly social episodes, the feminine superiority over the boys is especially noticeable. Test C, a typical cowboy "western" picture embroidered with a minor love story, takes a position between Tests A and B, from the point of view of the present discussion.

Less strikingly the following facts stand out: (a) Males seem to be superior in questions involving number (see questions Nos. 1 and 5 of the list above; No. 37, in which females excel, concerns an episode occurring in the midst of a highly social, somewhat romantic portion of the picture). (b) There is a hint that males are more interested in scenes about fighting (cf. Nos. 34 and 35 above), and that females are more susceptible to (at least, recall better) the tender, pathetic, "touching" episodes.

In some instances, a sex difference is shown on an item not because of its specific content, but because of the general setting in which the item has occurred. For example, female superiority in No. 6 may be attributed not to a greater interest of girls in baseball gloves, but to the fact that the matter concerning the baseball glove was brought out in the course of an intimate conversation between two lovers. The influence of differences in context may be systematically examined by averaging the percentage of correct replies by males and females for *all* items in a given context. This has been done in Table 5, which requires no comment.

The results of this section, emphasizing as they do the intimate influence of interest upon test performance, call into question certain studies of sex differences which have neglected to take interest factors specifically into account. (1; 2; 4; 12; 16, pp. 59-63; 20; 22).

TABLE 5
SEX DIFFERENCES AS DEPENDENT ON CONTEXT (TESTS A AND B)

Context	No. of questions	Percentage answering correctly			
		Adolescents		Adults	
		Male	Female	Male	Female
Women the center of action and interest, or directly implicated	7	29	46	54	61
Events directly involving a love interest	8	42	56	54	64
Events involving feeling and sentiment (other than a love interest)	6	26	33	37	53
Events at the party (the "shadow social")	8	52	64	66	68
Total	29	37	50	53	62
Events relating to baseball	7	30	30	49	33
Conflict situations (men the chief actors)	4	48	51	73	57
Total	11	39	41	61	45

SEX SIMILARITIES

The discussion to this point has emphasized differences between the sexes on individual test questions. The evidence shows clearly that such differences exist, and also indicates the locus and causes of the differences. Nevertheless, equally striking and important—perhaps more important—are the general sex similarities. Thus, the correlation between the percentage of correct answers by the adolescent males and the percentage of correct answers by the adolescent females, for all the questions in Test A, is no less than .93. Detailed data for this test, and the others, are given in Table 6.

The most reliable data in Table 6—that for Test A and Test B—show that the correlation between the sexes in the adolescent groups exceeds noticeably the correlation in the adult groups. At least a part of the greater discrepancy among the adults is due to the speed factor (pages 437-438).

The correlations in Table 6 have not been corrected for attenuation, principally because no precise measure is available for the reliability coefficient of the variables correlated. We have previously stated that the reliability of the motion-picture test total scores, for individual subjects, is .9 or over (footnote to Table 1). Here, however, we are dealing, not with total scores by individual subjects, but with the percentages of groups of adolescents and adults answering correctly on *individual items*. With small samples (say under 500),

TABLE 6
SEX SIMILARITIES IN THE PERCENTAGE ANSWERING INDIVIDUAL ITEMS
CORRECTLY

Test (1)	No. of ques- tions (2)	A d o l e s c e n t s						A d u l t s			
		Raw r (uncor- rected for at- tenua- tion) (3)	Mean of the percentages answering each item correctly		$S.D.$ _{dis.}		Raw r (uncor- rected for at- tenua- tion) (8)	Mean of the percentages answering each item correctly		$S.D.$ _{dis.}	
			Male (4)	Female (5)	Male (6)	Female (7)		Male (9)	Female (10)	Male (11)	Female (12)
A	40	.93	54	53	18	19	.82	65	57	17	18
B	48	.84	37	43	23	24	.79	53	51	23	22
$C_{immed.}$	20	.78	73	71	17	16	.78	82	78	13	16
C_{delay}	20	.51	55	57	16	17	.80	73	70	19	21

these percentages will, of course, vary sensibly from sample to sample. If two equivalent groups (M and N) were available, the correlation between the percentages passed by M and those passed by N would constitute a *reliability coefficient*, or a measure of sampling variation in terms of correlation. In the samples of the present study, the male and female adolescents are our most nearly equivalent groups. But the correlation between these two groups, for the various tests, has already been given in Column 3 of Table 6. The figures in Column 3 may therefore be regarded as *minimum* reliability coefficients, for each test, of the percentages passing each item—*minimum*, because the male and female adolescent groups are only incompletely equivalent. Of course, all the raw correlations between the sexes in Table 6 would be increased if corrected for attenuation; the degree of sex similarity shown in Table 6 is therefore to be considered conservative.

QUESTION TYPES

As previously stated, the questions in the three tests dealt with incidental episodes as well as essential "continuity," and with verbal (captional) as well as pictorial material. For the analysis to be presented in this section, the motion-picture content corresponding to each question was classified with regard to "significance" and to "verbalness." The questions were labeled X if significant to the development of the plot, Z if purely incidental, and Y if borderline (i.e., neither exactly essential nor yet incidental.) Similarly, the symbol I was assigned if the question was based on purely pictorial

material, 3 if based on purely captional material, and 2 if based on both pictorial and captional. Below are given illustrations of several combinations of type questions.

X 1 Significant, pictorial:

Why did Bert become angry during his talk with Molly?
(*His rival entered the house.*)

X 3 Significant, verbal:

When the telegram was read, Ben said (*"I'll try to do my best"*).

Z 1 Incidental, pictorial:

How many British were at the council? (*Four.*)

Z 3 Incidental, verbal:

Who was Governor of Virginia? (*Patrick Henry.*)

The classification of the questions was made by three psychologists who were thoroughly familiar with each film. The 108 items in Tests A, B, and C were fairly evenly divided among the several categories, except that there were relatively few items classified as 2 (based on both pictorial and captional material).

The questions now arise: (*a*) Is recall for any one of these types of questions significantly better than for any other? (*b*) Do sex differences in recall occur for any of these types? Table 7 presents the detailed data, in terms of weighted average percentages.

Weighting was theoretically desirable because items of the various types could not be evenly distributed throughout each test (due to the necessity of respecting the time sequence of the motion picture). Thus, a relatively large number of essential items appear in the last quarter of Test B; a relatively large number of captional items appear in the first quarter of Test C; etc. In order to minimize the effect of position-in-series, as well as the effect of the time limit under which the tests were administered, each test was divided into consecutive quarters; and the examples of each question type occurring in any quarter of a test paper were *averaged and treated as one*. For any particular question type, its recall value is given by the sum of the credits in each quarter, divided by 4.⁸ It may be noted that

⁸For example, in Test A, two questions of Type Y3 in the first quarter of the test were answered correctly by 68 and 76%, respectively, of the adolescent males: the score for this type for the first quarter of the test, therefore, is 72%. Similarly, the average score for this type, Y3, in the second quarter of Test A for the adolescent males is 55; for the third quarter, 75; for the fourth quarter, 32. The average recall value of Type

Y3, Test A, for adolescent males, is therefore $\frac{72+55+75+32}{4} = 58.5$.

TABLE 7
RECALL VALUES OF QUESTION TYPES

Line no. (1)	Question type (2)	Total no. of questions of type (3)	Adolescent recall values (percentages)										Adult recall values (percentages)																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
			Test A					Test B					Test C _{immed.}					Test C _{delay}					Test A					Test B					Test C _{immed.}					Test C _{delay}					Aver. Aver.																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
			M*	F*	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F

*"M" stands for *male*, "F" for *female*. The number of cases in each group is given in Table 1.

while the distribution of the various question types produced a theoretical requirement for weighting, the weighted values are actually closely similar to the unweighted averages.

Table 7 reveals slight fluctuations in the various question types within a given test. These fluctuations, however, are not consistent in the various tests. For example, there is no consistent tendency for significant questions (*X*) to be remembered better than incidental questions (*Z*),⁹ nor for pictorial items (*I*) to excel verbal items (*3*). Furthermore, there are no consistent sex differences among these various types of items. If comparisons are made, more minutely, in terms of *X1*, *X2*, *X3*, *Y1*, etc., the foregoing results are confirmed. It is clear, then, that the *general* characteristics of items (pictorial vs. verbal, significant vs. incidental) are relatively unimportant in producing either differences in difficulty, or differences in performance of the sexes. The latter kind of difference, at least, appears to depend almost wholly upon interest factors and specific selective appeal. It may seem strange that questions based on significant content are not better remembered than those based on content less essential, or even quite incidental. However, the incidental questions, in spite of their disconnected character, have the advantage of being concrete and specific, while the significant questions are sometimes more difficult because of the requirement of inference and an understanding of the film continuity. As an example of this, in Test B a quite irrelevant question ("How was the fence gate fastened?") was answered correctly by about 70% of both adolescents and adults; but a question of fundamental importance for the plot, and based on a much longer portion of screen narrative ("How did Ben show his feeling when he read the telegram?"), was answered correctly by only about 40% of the adults and 20% of the adolescents. Ability to answer any given item depends, among other factors, upon general intelligence, grasp of the plot, interest in the specific item, and interest in the general context in which the item occurs. While the first two of these factors may favor a higher score on the significant questions, the last two result in an individual's noticing and recalling those episodes of a motion picture which are related to his personal orientations; these are the episodes involving background, local color, comic relief, etc., which are often merely

⁹This is in spite of the fact that there is a clear and marked distinction, so far as significance to the plot is concerned, between the items marked *X* and those marked *Z*.

subsidiary and incidental. From this we may draw a corollary that, from the standpoint of concentration of interest and rapport with the audience, the most effective and best remembered sequences are, in general, those which are most fully "integrated," that is, those which are successful in including incidental elements in the essential pattern and movement of the scene.

Table 8 presents an analysis of sex similarities in terms of correlations. As in the case of Table 6, the degree of sex similarity indicated by the correlations is fairly high. Table 8 is to be read as follows: among adolescents the percentage of males answering correctly each question of Type *X1* correlates .71 with the percentage of females answering these questions correctly; among adults, the sex correlation is .86. As in Table 6, the correlations would be higher if corrected for attenuation; again the correlations are, on the average, higher for adolescents than for adults.

A COMPARISON OF ADOLESCENTS AND ADULTS

From the standpoint of the social and educational effects of motion pictures, it is important to consider the particular elements in pictures

TABLE 8
SEX SIMILARITIES IN TERMS OF CORRELATIONS

Question "type" (Tests A, B, and <i>C_{immed.}</i>)	No. of questions of specified "type"	Correlation (ρ) between sexes for	
		Adolescents	Adults
<i>X1</i>	10	.71	.86
<i>X2</i>	10	.94	.92
<i>X3</i>	17	.91	.92
<i>Y1</i>	12	.80	.51
<i>Y2</i>	13	.90	.76
<i>Y3</i>	10	.85	.77
<i>Z1</i>	23	.95	.95
<i>Z3</i>	13	.94	.92
Average		.875	.826
<i>A</i>	37	.87	.85
<i>B</i>	35	.88	.76
<i>C</i>	36	.93	.94
<i>1</i>	45	.82	.79
<i>2</i>	23	.93	.84
<i>3</i>	40	.92	.90
Average		.891	.847

which arouse the attention and interest of children. Of possible significance, also, for the psychology of testimony, is the comparison of adolescents and adults in range or fullness of report; the reliability of children as witnesses has often been investigated, but with widely varying conclusions (21, pp. 33-34).

The median age of the adolescents in the various tests is close to 14 years. The median age of the adults ranges from 25 to 31 in the case of the males; in the case of the females it is 34 for each test except C_{delay} , in which the median is 41. If we turn to Table 1, we may note that the inferiority of the adolescents in mean score is consistently maintained for each test and for both sexes, but the degree of inferiority is in no case very great. In terms of the variability of the adults, the male adolescents average about .6 *S.D.* below the male adults in the immediate tests; the female adolescents average about .3 *S.D.* below the female adults. This implies, of course, a marked degree of overlapping; as a general average, approximately 37% of the adolescents equal or excel the mean for the adults. In the tests delayed one week, the superiority of the adults is somewhat greater—about 1 *S.D.* for the males, and .6 *S.D.* for the females. These differences, in both immediate and delayed tests, are no doubt attributable chiefly to differences in intelligence between the adolescents and adults, and also between the male adults and the female adults (Table 3). In the one case in which an adolescent sample happens to equal an adult sample in Alpha score (females, Test A—see Table 3), the two samples are also substantially equal in recall score. We can find no support for McGeoch's theory of an *intelligence threshold* beyond which "there is no definite relationship between intelligence and ability to report" (11, p. 444). Within our broad range of intelligence (Alpha scores from 15 to 200), the regressions between intelligence and recall test score are linear. In terms of age groups, the mature adults surpass the children, exactly as normal children have been proved to surpass subnormals in both intelligence and capacity to report (11, 14).

In terms of percentages, we may say that, in the *average* test given immediately after a motion picture (with our sample and our procedure), adults are able to report correctly on about 65% of the items, while adolescents score an average of about 56%. In the case of Test C_{delay} , the introduction of the week's interval between exhibition of the picture and testing seems to strengthen the differences between adults and adolescents; for this test, the percentages of re-

TABLE 9
 ADOLESCENT-ADULT SIMILARITIES IN THE PERCENTAGE ANSWERING
 EACH ITEM CORRECTLY

Test	Rank order ρ , male adolescents \times male adults	Rank order ρ , female adolescents \times female adults
A	.90	.84
B	.82	.84
<i>C_{immed.}</i>	.90	.70
<i>C_{delay}</i>	.70	.82

tention are 75 and 58, respectively. For a comparable group, the percentages of retention for Test *C_{immediate}* are about 89 and 77. The drop from the immediate retention to the retention after a week's delay appears, therefore, to be relatively small; the delayed retention is in fact almost 90% of the original retention for the adults, and about 75% of the original retention for the adolescents. These are, after all, decidedly high figures for retention, and they suggest the marked "staying power" of material acquired in the observation of motion pictures.

We may next inquire whether the differences between adolescents and adults are attributable specifically to certain elements in pictures; we have seen that sex differences, which are small in the aggregate scores, are appreciable in terms of individual items which appeal more to the interests of one sex than the other. Similarly, are there in our pictures characteristically "adult" and "adolescent" items? The correlation between the adolescent and the adult performance in the questions of each test is given below in Table 9. (Table 9 is to be read similarly to Tables 6 and 8, already presented.)

The correlations between adolescent and adult performances in Table 9 are about as high as those between the sexes in Tables 6 and 8.¹⁰ But, whereas a detailed examination of sex differences in individual questions proved fruitful, an equally systematic and detailed examination of the adolescent-adult differences on individual questions fails to afford more than one or two doubtful clues. The

¹⁰Note that the correlation in Table 9 between male adolescents and adults is fully as high as between female adolescents and adults. This fails to support the belief that the relatively rapid rate of sexual maturity among females minimizes the differentiation of interests within this sex, whereas the slower rate of sexual maturity among males heightens such differentiation.

individual questions in which the adolescents approach (or even excel) the adults seem to be those based on content which is only very briefly exposed in the motion picture, and which is of an incidental, detailed, or exact (numerical) nature. The adolescents give some evidence of falling exceptionally far behind the adults in cases where the correct answer involves an inference from the content of the motion picture, particularly if this inference must be derived from a social situation. This is in agreement with the view that the "social intelligence" of adolescents lags slightly behind their abstract intelligence. The differences just given are none too firmly established, and are presented only tentatively. Doubtless, with pictures representing a wider range of content, or with samples involving a wider range of age, clearer divergencies would be obtained.

It has already been indicated that, on the individual items, the sex differences among adolescents appear somewhat smaller than among adults. Some interest attaches to the question as to whether or not a sex difference in the adolescent group (on a specific item) forecasts a similar sex difference in the adult group. For each motion-picture test, we have taken the difference between the percentage of adolescent males and adolescent females passing each item, and correlated this with the difference between the percentage of adult males and adult females on the corresponding items. The results for each test¹¹ are presented in Table 10, and indicate little direct relationship between adolescent and adult sex differences. Had the correlation been restricted to those items in each test for which reliable (98/100) adolescent or adult sex differences were obtained, the results would be virtually the same; thus, for the 25 items showing such a degree of reliability in Test B (see pages 442-444), the correlation (ρ) between adult and adolescent sex differences is .36. If, as has been affirmed, sex differences are due chiefly to differences in interest, it would appear that the interests of boys do not differ from those of girls in a manner which directly predicts the differences between men and women. In other words, sex differences themselves are subject to differentiation extending through and beyond the adolescent period. This is in accord with the theory of specificity of sex differences and of their dependence upon developmental changes in specific interests. Doubtless, certain adolescent sex differences wane in maturity, whereas, others increase.

¹¹Test C_{delay} is omitted because the number of cases is too small to give adequate reliability to the differences between the obtained percentages.

TABLE 10
CORRELATION BETWEEN ADOLESCENT SEX DIFFERENCES AND ADULT
SEX DIFFERENCES

Test	No. of questions	Correlation (rank order ρ)
A	40	.30
B	48	.22
C _{immed.}	20	.26

SUMMARY AND CONCLUSIONS

1. Motion pictures present many technical advantages in the study of testimony and other phases of memory. These advantages include a wide range of subject-matter, the provision of incentives for unselected samples, and precision of experimental control.

2. The present report is based on testimony for the content of three motion pictures. Motion-picture test records were obtained for 353 adolescents (10-17 years) and 393 adults (18-54 years).

3. When allowances are made for minor inequalities of sampling, no sex differences are observable in either the central tendency or variability of total test scores.

4. An item analysis of the 108 individual test elements in the three motion-picture tests reveals a considerable number of individual items with a sex difference having a reliability of 98/100 or greater.

5. The items of each test were classified with respect to their captional or pictorial basis, and with respect to their significance in the film continuity. The various classifications yield no difference in average difficulty, and no systematic differences between the sexes.

6. Since sex differences on individual items are not to be explained on the basis of the general characteristics of the items, attention must be given to specific factors of interest or selective appeal.

7. A measure of sex resemblances in the individual test items is secured by correlating the percentage obtained by males on each item with the percentage obtained by females. These sex similarity coefficients cluster about .8 for both adolescents and adults. They indicate that the difficulty of individual items tends usually to be similar for the two sexes.

8. Comparisons are also made between adolescents and adults, without special regard to sex. In terms of total test scores, and using

the *S.D.* of the adults as a unit, the male adolescents average about .6 *S.D.* below the male adults in the immediate tests, and about 1 *S.D.* below in tests delayed one week. Among the females, the adolescents average .3 *S.D.* lower in the immediate tests and .6 *S.D.* lower in the delayed tests. Such differences, of course, imply considerable overlapping between the adolescent and the adult groups. The differences are attributable chiefly to the correlation of intelligence with chronological age and with motion-picture test scores.

9. When the performance of adolescents in the individual test items is compared with that of adults, no certain systematic causes of differences (analogous to those producing sex differences) are found. The correlations between the adolescents' and adults' performances on individual items (each sex taken separately) range from .7 to .9.

10. When adolescent sex differences on individual items are correlated with corresponding adult sex differences, coefficients of from .2 to .3 are obtained (uncorrected for attenuation). This indicates, at least for the material in question, that adult sex differences cannot be forecast directly from the specific sex differences found at an earlier age.

11. In terms of percentages, in the average test given immediately after a motion picture, adolescents are able to report correctly on about 56% of the items, while adults score an average of 65%. After an interval of one week the retention scores remain higher than expected: adolescents score about 75% of what they would have obtained in an immediate test, and adults about 90%.

12. Prior studies reporting sex differences in memory scores cannot be assumed to have demonstrated differences in such a generic trait as "power to recall"; they may depend merely upon the averaging of many specific responses, and ultimately upon the differential interest values in the particular assortment of items which have been brought together. The present study indicates also the possible importance of sampling inequalities, with respect to the factors of intelligence and of speed in test performance.

REFERENCES

1. ALLEN, C. N. Studies in sex differences. *Psychol. Bull.*, 1927, **24**, 294-304.
2. ———. Recent studies in sex differences. *Psychol. Bull.*, 1930, **27**, 394-407.

3. BASSETT, S. J. Sex differences in history retention. *School & Soc.*, 1929, **29**, 397-398.
4. BORING, E. G. Capacity to report upon motion pictures as conditioned by sex and age. *J. Crim. Law & Criminol.*, 1916, **6**, 820-834.
5. CONRAD, H. S., & JONES, H. E. Psychological studies of motion pictures: III. Fidelity of report as a measure of adult intelligence. *Univ. Calif. Publ. Psychol.*, 1929, **3**, 245-276.
6. ———. Psychological studies of motion pictures: IV. The technique of mental test surveys among adults. *Univ. Calif. Publ. Psychol.*, 1929, **3**, 277-283.
7. GARRETT, H. E. Statistics in psychology and education. New York: Longmans, Green, 1926. Pp. 330.
8. JONES, H. E. A first study of parent-child resemblance in intelligence. *27th Yrbk. Nat. Soc. Stud. Educ.*, 1928, Pt. 1, 61-72.
9. ———. Emotional factors in learning. *J. Gen. Psychol.*, 1929, **2**, 263-272.
10. JONES, H. E., CONRAD, H. S., & HORN, A. Psychological studies of motion pictures: II. Observation and recall as a function of age. *Univ. Calif. Publ. Psychol.*, 1928, **3**, 225-243.
11. MCGEOCH, J. A. The fidelity of report of normal and subnormal children. *Amer. J. Psychol.*, 1925, **36**, 434-445.
12. ———. The influence of sex and age upon the ability to report. *Amer. J. Psychol.*, 1928, **40**, 458-466.
13. MYERS, G. C. A study in incidental memory. *Arch. Psychol.*, 1913, No. 26. Pp. iii+108.
14. PEAR, T. H., & WYATT, S. The testimony of normal and mentally defective children. *Brit. J. Psychol.*, 1914, **6**, 387-419.
15. PYLE, W. H. The relation of sex differences to the kind of material used. *J. Educ. Psychol.*, 1925, **6**, 261-264.
16. SCHRENK, J. Einführung in die Psychologie der Aussage. Leipzig: Quelle & Meyer, 1922. Pp. viii+110.
17. SHELLOW, S. M. Individual differences in incidental memory. *Arch. Psychol.*, 1923, No. 64. Pp. 77.
18. SMITH, W. W. The measurement of emotion. London: Kegan Paul; New York: Harcourt, Brace, 1922. Pp. 184.
19. SUGISAKE, Y., & BROWN, W. The correlation between sex of observers and sex of pictures recognized. *J. Exper. Psychol.*, 1916, **1**, 351-354.
20. THOMPSON, H. B. The mental traits of sex. Chicago: Univ. Chicago Press, 1903. Pp. vii+188.
21. WHIPPLE, G. M. Manual of mental and physical tests. (Vol. II.) Baltimore: Warwick & York, 1921. Pp. 336.
22. WOOLLEY, H. T. The psychology of sex. *Psychol. Bull.*, 1914, **9**, 353-379.

University of California
Berkeley, California

ÉTUDES PSYCHOLOGIQUES DE FILMS: V. DIFFÉRENCES DE SEXE CHEZ LES ADOLESCENTS ET LES ADULTES DANS LE RAPPEL IMMÉDIAT ET RETARDÉ

(Résumé)

On ne peut dire que les études antérieures rapportant les différences de sexe dans les tests de mémoire ont montré des différences dans un trait générique tel que "la capacité de rappel," car leurs résultats peuvent dépendre seulement de la moyenne de plusieurs réponses spécifiques, et enfin des valeurs différentielles d'intérêt dans les choses spéciales employées. Dans cette enquête, on a obtenu des témoignages sur le contenu de trois films du cinéma, de 351 adolescents (10-17 ans) et 393 adultes (18-54 ans). Les films ont inclus un total de 11 "reels" et ont été testés par 108 questions de test avec une grande variété de sujets. Dans le test moyen subi (non annoncé) immédiatement après un film, les adolescents ont pu rapporter correctement dans environ 56% des questions, tandis que les adultes ont obtenu une moyenne de 65%. Après un intervalle de huit jours (avec des choses non testées antérieurement) les adolescents ont obtenu environ 75% de ce qu'ils auraient obtenu dans un test immédiat, et les adultes environ 90%. Les résultats totaux n'ont révélé aucunes différences significantes de sexe. La difficulté des questions individuelles a tendé à varier de même manière pour les deux sexes, comme le montrent des corrélations de l'ordre de 0,8 entre les pourcentages des questions pour les mâles et pour les femelles. Plusieurs questions individuelles cependant ont montré des différences bien marquées entre les pourcentages des mâles et ceux des femelles; celles-ci n'ont pu être attribuées aux traits généraux des questions, mais aux facteurs spécifiques d'intérêt ou aux facteurs choisis d'intérêt ou d'attrait choisis.

CONRAD ET JONES

PSYCHOLOGISCHE UNTERSUCHUNGEN AN WECHSELBILDERN: V. GESCHLECHTSUNTERSCHIEDE IN BEZUG AUF UNMITTEL- BARE VERZÖGERTE WIEDERHOLUNG BEI JUGENDLICHEN UND ERWACHSENEN

(Referat)

Es kann nicht angenommen werden, dass frühere Untersuchungen, die über Geschlechtsunterschiede bei Prüfungen des Gedächtnisses Bericht erstatten, die Anwesenheit von Unterschieden bei einer so allgemeinen Eigenschaft wie sie das "Wiederholungsvermögen" ("Power to recall") darstellt, erwiesen haben, denn die Ergebnisse dieser Untersuchungen ruhen vielleicht bloss auf dem durchschnittlichen Betrag vieler einzelnen Erwiderungen, und zuletzt auf Unterschieden in der Interesse der einzelnen Punkte (items) für die Versuchspersonen. In der vorliegenden Untersuchung erhielt man von 351 Jugendlichen (10 bis 17 Jahre alt) und 393 Erwachsenen (18 bis 54 Jahre alt) Aussagen über den Inhalt von drei Filmen. Man untersuchte im Ganzen 11 Rollen Film, worüber 108 Testfragen, sich über eine grosse Auswahl von Gegenständen erstreckend, gestellt wurden. Durchschnittlich gelang es den Jugendlichen, 56% der sofort nach Besichtigung des Filmes (unangemeldet) gestellten Testfragen richtig zu beantworten, während Erwachsene durchschnittlich auf 65% richtig antworteten. Nach einer wochen-

langen Pause erzielten die Jugendlichen (an noch nicht geprüfem material) ungefähr 75%, und die Erwachsenen ungefähr 90% so viel, wie sie in einer unmittelbar auf die Darstellung folgenden Prüfung erreicht hätten. Die Gesamtzahlen erwiesen keine bezeichnenden Geschlechtsunterschiede. Die Schwierigkeit der einzelnen Punkte (items) erwies die Tendenz, bei beiden Geschlechtern auf gleiche Weise zu variieren. Das lies sich daraus sehen, dass Korrelationen von ungefähr .80 zwischen den richtig beantworteten Fragen bei männlichen und denen bei weiblichen Versuchspersonen bestanden. Bei vielen einzelnen Punkten, jedoch, zeigten sich zuverlässige Unterschiede zwischen den Prozentzahlen bei den männlichen und denen bei weiblichen Vpp. Diese Unterschiede liessen sich nicht auf allgemeine Eigenschaften der einzelnen Punkte zurückführen, sondern auf spezifische Elemente der Interesse oder der besonderen Anlockung (selective appeal).

CONRAD UND JONES

THE VERSATILITY OF GENIUS*¹

From the Psychological Laboratories of Stanford University

RALPH K. WHITE

PURPOSE

The purpose of this study is twofold: (*a*) to estimate the versatility of three hundred eminent men, as an indication of the extent to which specialization is favorable or unfavorable to the attainment of eminence; and (*b*) to discover what kinds of special ability are associated with certain kinds of genius, as an indication of the vocational types to be kept in mind in the education and guidance of gifted children.

SUBJECTS AND METHOD

The subjects used were the three hundred eminent men studied by Dr. C. M. Cox (1) in her book, *The Early Mental Traits of Three Hundred Geniuses*, and the data on their special abilities are drawn almost entirely from the material collected in the preparation of that book. For a complete description of the method of selection, the composition of the group, and the nature of the data, the reader is referred to that volume. The criterion of inclusion was, almost of necessity, eminence or fame rather than ability estimated directly as such. Only individuals born after 1450 A.D. and before 1850 A.D. were considered.

Some three thousand volumes were consulted by the compilers of the original data, who were instructed to make note (among other things) of all significant facts showing special ability or deficiency along any line. The resulting collection of condensed information, on file at Stanford University, amounts to an average of about twenty typewritten pages per individual included. Because of the extensiveness of the material, only a very much condensed account of each individual case study, chiefly drawn from the material which had

*Recommended by Lewis M. Terman, accepted for publication by Carl Murchison of the Editorial Board, and received in the Editorial Office, October 12, 1930.

¹The study was initiated by Dr. Lewis M. Terman, and carried out under his direction. The writer is extremely grateful to Dr. Catherine Cox Miles and Dr. Calvin P. Stone for criticism and suggestions.

the most direct bearing on the estimation of the IQ, was published in the form of case studies (1, Part II).

The writer derived most of his material from the manuscript biographies; that is, from the twenty-page biographies, and not from the far more condensed case studies published in Dr. Cox's book. He transcribed or paraphrased, on a separate sheet of paper for each man, all items of information there collected in regard to special abilities or deficiencies. The information on versatility thus collected was then evaluated by two independent raters, Mrs. Ruth H. Thomson (formerly Mrs. Ruth Haines Livesay) and the writer. Mrs. Thomson, under Dr. Cox's direction, had collected the material for about one third of the original case studies, and was familiar with the general plan of the work. Both raters used the original files freely in supplementing the selected items on the "versatility sheets."

About ten per cent of the scores were later modified, and some new ones added, by Dr. L. M. Terman, and by Dr. E. M. Hulme and Dr. R. H. Lutz of the Stanford History Department, who very kindly looked over the records of men about whom they possessed special information. The scores given by them were averaged with those of the writer. (In computing reliability coefficients their ratings were simply substituted for those of Mrs. Thomson, in the few cases where such substitution was necessary.)

It was found that the data could be conveniently classified under 23 fields of ability, more or less arbitrarily defined as follows:

1. *Administration*: executive work in religious, educational, or scientific organizations. Business, politics, and warfare are not included.

2. *Art*: drawing, painting, sculpture, architecture, or unusual appreciation.

3. *Business*: money-making ability, or business administration. Negative scores denote, in some instances, general absentmindedness or impracticality.

4. *Conversation*: the non-scholastic, purely human side of conversation. "Charm."

5. *Drama*: writing or unusual appreciation.

6. *Handwork*: non-creative manual ability.

7. *History*: history and biography.

8. *Humor*: wit and humor shown in conversation or inferred from writings, the latter being discounted as less spontaneous.

9. *Invention*: creation of new mechanical devices.

10. *Languages*: foreign languages, ancient and modern.
11. *Law*: theory or practice of law.
12. *Mathematics*: pure mathematics, astronomy, engineering, surveying, navigation.
13. *Medicine*: medical practice or knowledge (research being included under "science").
14. *Music*: composing, performing, or unusual appreciation.
15. *Non-fictional prose*: essays, criticism, letters, controversial writing, journalism, or an excellent literary style in science, philosophy, etc.
16. *Novels*: novels or short stories; writing, story-telling, or unusual appreciation.
17. *Philosophy*: epistemology, ethics, theology, and theoretical psychology.
18. *Poetry*: verse of any kind; writing or unusual appreciation.
19. *Politics*: participation in politics or diplomacy; not political theory.
20. *Public speaking*: political speaking, lecturing, preaching, teaching (in some cases), acting.
21. *Science*: the relatively non-mathematical sciences; not astronomy.
22. *Social theory*: political, economic, and educational theory.
23. *Warfare*: military and naval activity; exploration.

Each man was given a rating in each of these 23 special abilities except the one in which his own primary achievement was assumed to lie. The rating was done on a subjective scale, with scores running from minus 5 to plus 5, 0 representing the assumed ability of the average college graduate of today in the trait in question.

"Versatility" was defined in terms of ability; but, to a certain extent, interest was taken as an indication of ability, and activity as an indication of interest. For example, in the single field of musical ability, our only information about the astronomer Kepler was that he "began to learn music at the age of eight."² He was given a score of 1 by both raters, because both believed that this activity indicated slightly more than a fifty-fifty chance that Kepler possessed musical

²References are not given for any of the numerous but fragmentary quotations scattered throughout this paper. The relative reliability of sources has, however, been definitely considered in the evaluation of data. A bibliography for each man is given by Cox (1). The source of any specific quotation will be furnished gladly, on request.

ability slightly greater than that which the average American college graduate would have had at the same age. Though positive, this piece of information is so meager as to be almost negligible. A score of 5, or the highest possible (since no scores were given in music to men who achieved eminence primarily in music), was granted to the astronomer Herschel, who played the violin at four, was later an oboist, bandmaster, manager of concerts, and organist, and who, at the ages of 21 and 22, wrote 12 symphonies. An attempt was then made to divide the difference between Kepler and Herschel into four equally-noticeable intervals, and individuals falling between were scored accordingly. For instance, a score of 2 was given to the poet Longfellow, who was fond of playing the flute. A score of 3 was granted to Washington Irving because, at the age of ten, he had "a love of music which became later in life a passion." The philologist Wolf earned a score of 4 by studying music (singing and clavier playing) at the age of three, becoming accomplished later in string and wind instruments, and composing new airs between the ages of fourteen and eighteen. In each field, mere interest or activity was rarely given a score above 2; definite evidence of ability without originality was rarely scored above 3; and the scores of 4 and 5 were reserved for creative achievement of a rather high order. Concrete evidence of achievement coming from a reliable source was also, of course, scored higher than subjective estimates or information coming from questionable sources.

Individuals about whom there was no information (there were 213 in the field of music, for example) were given scores of 0. On the other side of the zero line, the same method was used. For example, Immanuel Kant was given a score of -2 because he "disregarded music." The lowest score in music (or any other field) was given to Alexander Humboldt, who considered music a "social calamity." One rater gave him -5 , the other -4 . Here, as always when the ratings differed (as they did in about 40% of the cases), the average, or -4.5 , was taken as the final score.

Reliability coefficients, measuring correspondence between the two raters' estimates, were computed for only two abilities; but, since these agreed fairly closely and seemed by inspection to be representative of the whole, it was not considered necessary to go further. The coefficients (not stepped up in any way) were, for languages, $.75 \pm .02$, and for politics, $.78 \pm .03$.

This surprisingly high reliability does not imply, of course, equally

high validity. The two raters were using almost exactly the same information, limited in scope, and usually quite specific. The fact that they tend to interpret it in the same way does not remedy the fact that the original biographical information depends in many cases upon subjective estimates made by the associates of the individual, who were untrained observers, or even on hearsay evidence. (See criticism, page 480).

As a partial check on validity, three individuals were investigated with considerably more thoroughness than any of the others. These were the chemist Davy, the philosopher Spinoza, and the statesman Webster. They were selected in order to be as nearly as possible representative of the group as a whole in number of versatility-points given previously (an extremely rough measure, not used anywhere else in this study), in amount of additional material available, and in eminence as indicated by Cattell's ranking. An average of 5.7 volumes were consulted for each of the three, and, in the light of the new information, each score previously given was re-examined. Only one person, unfortunately, was able to do this re-examination (the writer). As a result of it, the number of abilities on which there was some information was increased 30%. The average excess of positive over negative scores was increased; that is, these men seemed to be slightly more versatile after the follow-up than they had seemed before. The excess of positive scores decreased from 18.5 to 18 in the case of Davy, increased from 18.5 to 23.5 in the case of Spinoza, and remained unchanged at 20 in the case of Webster.

The scores obtained were recorded on a large chart, and the average score in each kind of ability was computed separately for 12 classes of geniuses. In at least 95% of the cases the classification of the subjects was that given by encyclopedias, but in a few the writer used his own judgment in order to bring the classification into conformity with the definitions of special abilities given above, on pages 461-462. The number of individuals included in each class was as shown in Table 1.

These types of genius were then ranked for each ability, and each type compared with the mean for the whole three hundred. Since there are no norms derived from comparable or control groups, this comparison with the mean of the group as a whole seemed to be the only method that could be used in constructing a trait profile, in terms of special abilities, for each type.

TABLE 1

Class	Number
Statesmen	57
Novelists and dramatists	32
Soldiers and explorers	32
Philosophers and social theorists	31
Poets	26
Scientists	26
Religious leaders	22
Historians and philologists	19
Non-fictional prose writers	19
Artists	13
Mathematicians and astronomers	12
Musicians	11
Total	300

RESULTS

General Versatility

Positive vs. negative scores. There was an overwhelming preponderance of positive as compared with negative scores in ratings of the separate abilities (Table 2).

TABLE 2

	Number	Percentage of total
Scores above zero	2015	30.5
Scores below zero	141	2.1
Zeros (no information)	4450	67.4
Total	6606	100.0

The ratio of positive to negative scores is approximately 14.3 to 1. There was almost complete agreement of the two raters on this point; the ratio of positive to negative in the original ratings by Mrs. Thomson was approximately 14.5, and, in the original ratings by the writer, 14.4. Only two men (Ney and Rembrandt) had more negative than positive scores, and only three others (Palestrina, Mozart, and Van Dyke) had an equal number of positive and negative. That is, of 300 subjects, 295 had an excess of positive scores.

At least 90% of the additions made by Professors Terman, Hulme, and Lutz were also on the positive side. In the additional information obtained about Davy, Spinoza, and Webster, positive corrections predominated, but to a less extent. Eight positive and three negative corrections were made; these resulted in 12.5 additional positive points, and 7 additional negative points.

The objection may be raised that the zero point has been set too low, i.e., that both raters have too low an opinion of the capacities of the "average college graduate." Though the raters were familiar with college communities in Connecticut, Michigan, California, and the Hawaiian Islands, and though they arrived independently at almost exactly the same conclusion (a positive-negative ratio of 14.5 in one case and 14.4 in the other), they are quite willing to admit that this criticism may be valid. The reader is urged to judge for himself on the basis of a few examples of the kind of data given a score of 1, or the lowest possible positive score:

Darwin, to whom "algebra was repugnant," but who "took great pleasure in Euclid," was scored 1 in mathematics. (He may easily have deserved a score of 3 or 4, but our data were not full enough to warrant more than 1.) Byron, who read Blackstone and Montesquieu for pleasure, was scored 1 in law. The poet Klopstock, said to possess "personal attractiveness and sociability," was scored 1 in conversation. Admiral Farragut, whose "graphic account of what he saw in the various cities (age 15) would put many an older traveler to shame," was scored 1 in non-fictional prose. The novelist Dumas, who "picked up a knowledge of drugs and of anatomy," was scored 1 in medicine. Victor Hugo, who "obtained some distinction" in physics at school, was scored 1 in science. The figures 2015 and 141, then, mean that there were about two thousand cases in which the available evidence was as favorable as this or more favorable, and about one-fourteenth as many in which the available evidence was distinctly less favorable than the examples just cited.

It should be noted, too, that while the average negative score was very small, the average positive score was distinctly larger and hence more reliable than the examples just cited. The distribution, in percentage, of all the positive or negative scores given, is shown in Table 3. The same distribution is presented graphically in Figure 1. The mean positive score was 2.43, and the mean negative score, -1.40. The mean score of both classes together, disregarding zeros, was 2.18.

TABLE 3

Score	Percentage of total
5	4.1
4.5	4.1
4	6.0
3.5	7.0
3	13.3
2.5	12.3
2	19.5
1.5	10.0
1	13.1
0.5	4.1
-0.5	.9
-1	2.4
-1.5	1.5
-2	1.1
-2.5	.3
-3	.2
-3.5	.0
-4	.0
-4.5	.1
-5	.0

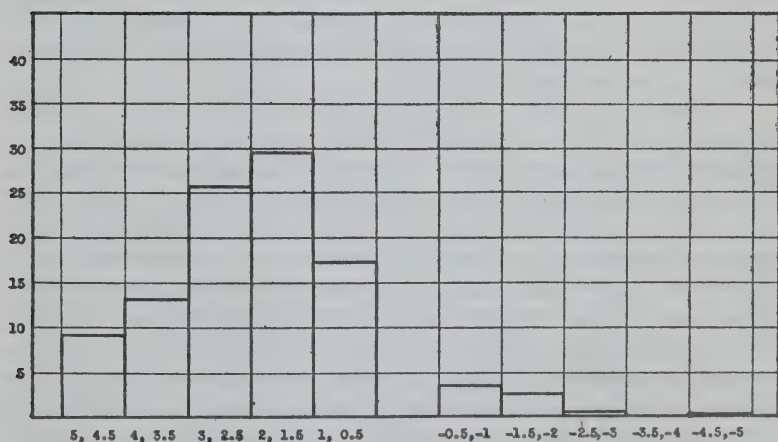


FIGURE 1
DISTRIBUTION OF 2015 POSITIVE AND 141 NEGATIVE SCORES

It is clear that the scores form a rather normal distribution, with the exception that the upper extreme is cut short, and that there is a gap between 0.5 and -0.5 representing the 4450 abilities about which there was no information at hand.

Typical positive scores would be the following: John Quincy Adams, who acquired a "passion for versification" at the age of 14, was scored 2 in poetry. The poet Béranger, who "outlined some comedies" at the age of 21, was scored 2 in drama. Edmund Burke, who "turned back to logic and metaphysics" at the age of 16, indulging in what he later called his "furor logicus," and who described philosophy as "queen of arts and daughter of heaven," was scored 2 in philosophy. The philosopher Campanella, who at the ages of 15 to 22 "minutely compared the Greek, Latin, and Arabian commentators" on Aristotle, was given a score of 3 in languages. Other examples will be found in the case notes given as an appendix to this paper.

On the other hand, typical negative scores would be the following: Victor Hugo, who confessed that philosophy did not interest him, was scored -1 in philosophy. The scientist Hunter, whose "lectures were not always clear," and who "preferred not to lecture," was scored -1 in public speaking. And, because his manners were "impatient, blunt, and unceremonious," and because he was "usually taciturn," he was scored -2 in conversation. The reader will probably agree that these typical negative scores seem to be more tentative and unreliable than the typical positive scores cited above, which outnumber them 14 to 1.

Interest vs. ability. The objection may be raised (see criticism on page 481) that no adequate distinction has been made between versatile abilities and versatile interests. An attempt was made to check up on this point, as far as possible, by going over all of the data once more and classifying each item as either an "interest item" (interest alone mentioned in the records) or an "ability item" (ability, or both interest and ability, mentioned in the records). It was found that 515, or 24%, belonged to the former class, and 1568, or 76%, to the latter. It was also found that the ratio of the number of positive scores to the number of negative scores was 6.6 to 1 in the interest items, and 21.5 to 1 in the ability items. In other words, the men studied seem to stand out from the average more in versatile ability than in versatile interest.

This accounts in large part for the fact that positive scores were

higher on the average (2.4) than negative scores (—1.4). Interest items were deliberately given lower scores than ability items by both raters. Nearly all of the positive scores above 2 were based on evidence of real ability; but, since evidence of real disability was very scanty, there were very few negative scores above 2 to counterbalance them.

The age factor. Many of these scores are based on information which applies directly only to the early years of a man's life. Robert Burns was given a score of 1 in drama because at the age of 17 he "sketched the outlines" of a tragedy. Others are based on information which applies only to the later years; Copernicus, for instance, was scored 2 in social theory because at the age of 49 he wrote a "memorandum on the confused state of the currency." Dr. Cox's data, collected primarily to throw light on the "early mental traits" of three hundred geniuses, were necessarily incomplete for the later years. It was suggested that perhaps the versatility found in this study is a temporary thing, a kind of "exploratory behavior" which tends to pass away as the man grows older and settles down to his main business in life.

To check up on this hypothesis, so far as it was possible to do so from the data at hand, the first 150 men in the alphabetical series were reviewed by the writer, and their positive scores classified as follows:

301, or 32%, apply directly to ages before 26 only.

214, or 22%, apply directly to ages after 26 only.

441, or 46%, probably or certainly apply to both periods.

SUMMARY

In other words, although clearly inadequate with respect to the age factor, the data are not very conspicuously onesided in their emphasis on the early years.

Relation to achievement. The writer also reviewed the first 150 men and checked each score which, in his opinion, represented an ability that contributed directly to achievement in the individual's major field. For example, Gibbon's score of 2 in languages, Burke's score of 3 in history, Drake's score of 4 in mathematics (navigation), and Heine's score of 5 in humor, were checked as probable examples of this type. Thirty per cent of the scores were checked off on this basis.

The remaining 70% were then subdivided again into abilities that,

in the writer's judgment, might have taken time away from the occupation by which eminence was achieved, and those that apparently could not. The arbitrary criterion used was 50 hours; for instance, when Burns sketched the outlines of a tragedy, it probably took up less than 50 hours of his time, but when Goethe is given a score of 5 in science because of volumes on comparative anatomy, geology, and the psychology of color, it undoubtedly represents much more than 50 hours. If he had not spent so much time on science, he might have written more novels or poetry. Fifty hours is nothing in comparison with a lifetime, of course, but the criterion was intentionally set low in order to give minimum figures for the non-time-consuming group. Even with this low criterion, it was found that about 4/7 of the 70%, or 40% of the total, belong with Burns' tragedy in the non-time-consuming category; and 3/7 of the 70%, or 30% of the total, belong with Goethe's science in the more time-consuming category.

To express the same facts more concretely, it can be said that the average genius is represented in our data as taking part in:

About 2.0 activities which contribute to eminence.

About 2.7 activities which probably neither contribute to nor hinder the attainment of eminence.

About 2.0 activities which *possibly* hinder the attainment of eminence.

These are, perhaps, the roughest, most subjective, least reliable figures in this study. They are presented, not as valuable data in themselves, but as a warning against the uncritical assumption that all positive scores should be given equal weight, or that versatility is often carried to the point where it actually interferes with achievement in the primary field.

Relative frequency of abilities. The distribution of the scores in each special ability was as shown in Table 4.

This rank order, of course, has little meaning. The fact that 17 received positive scores in invention is probably much more significant than the fact that 67 received positive scores in music. Three hundred college graduates picked at random might easily include 67 who were very musical, but would hardly be expected to furnish 17 inventors. Each figure should also be interpreted in terms of what was expected of individuals of the same social status and at the same period of history. For instance, the low position of science as compared with languages and philosophy may simply reflect the lack of recognition

TABLE 4

	Positive	Negative	No information	Positive, percentage of total
Conversation	189	29	82	63.0
Languages	175	7	113	61.7
Poetry	156	2	116	56.9
Philosophy	136	5	150	46.4
Non-fictional prose	119	0	162	42.3
Social Theory	113	1	164	40.7
History	115	1	171	40.1
Science	106	4	164	38.7
Humor	109	10	181	36.3
Public speaking	104	7	179	35.8
Politics	80	0	163	32.9
Mathematics	92	25	171	31.9
Drama	76	0	210	26.6
Art	67	2	218	23.3
Music	67	9	213	23.2
Law	66	9	224	22.1
Administration	62	4	232	20.8
Novels	43	2	235	15.4
Medicine	37	2	235	15.4
Business	35	19	246	11.7
Handwork	28	3	269	9.3
Warfare	23	0	245	8.7
Invention	17	0	281	6.0

given to science in education before the second half of the nineteenth century; and the low rank of business as compared with politics, strange as it seems today, may simply reflect the fact that until recently politics was generally held to be more "respectable" than business.

Ranking of types in general versatility. To obtain anything like a true measure of "general versatility" it would be necessary to weight each item according to the number and complexity of the component abilities included under it, according to its overlapping with other abilities taken account of, according to its lack of direct relationship with the main field of achievement, and according to its frequency in the social group to which the individual belonged. Since few if any of these corrections are feasible in the present state of our knowledge,

none of them have been applied. The simple average of the number of positive and negative scores has been computed, and the types ranked accordingly. The method is extremely crude, and cannot be relied upon except when differences are very well marked.

TABLE 5

Type	Av. no. of positive scores per individual	Av. no. of negative scores per individual	Difference
Non-fictional prose writers	8.2	.6	7.6
Statesmen	7.9	.5	7.4
Philosophers	7.8	.5	7.3
Scholars	7.4	.6	6.8
Religious leaders	7.1	.4	6.7
Scientists	7.2	.5	6.7
Poets	7.4	.7	6.7
Mathematicians	6.7	.1	6.6
Novelists and dramatists	7.1 \pm .3	.6 \pm .1	6.5 \pm .3
Soldiers	4.7 \pm .2	.4 \pm .1	4.3 \pm .2
Artists	4.2	.2	4.0
Musicians	3.3	.6	2.7

Probable errors have been computed for only two types, but the others are of the same order. It will be observed that differences within the first nine groups are not significant, but that the last three seem to form a group by themselves. The difference between novelists and soldiers, divided by its standard error, is 3.8; and the difference between musicians and any one of the first nine groups is correspondingly more dependable.

The reader is urged to interpret these figures in the light of his own interpretation of the more specific facts on pages 473-474. Except in verbal abilities, which probably made up an undue proportion of the 23 abilities considered, the soldiers and artists do not seem unversatile. The musicians, on the other hand, seem relatively weak not only in verbal abilities but also in all non-verbal abilities except art and humor.

Criteria of Vocational Types. The fact that a certain type of genius tends to possess a certain kind of special ability may be explained in at least four ways. It may indicate that the special ability is of direct value for the attainment of eminence in the major field, as Bacon's scientific interests contributed directly to the value of his philosophical writings. It may indicate a fundamental similarity

of mental processes or mental capacities, as Leibnitz's mental processes while working at mathematics may have been similar to his mental processes while working at philosophy. It may be that the temperament which is *interested* in one thing is attracted by similar elements in another; Madison's great interest in history would seem natural even though, possibly, it did not make him a much better statesman. Or, in some cases, the correspondence may be due to purely irrelevant factors, such as the nature of the particular educational system to which the individual was subjected. One should step very carefully, therefore, when attempting to formulate "vocational types" from such data as have been collected in this study. Each one of the correspondences listed below should be examined in the light of the four possible explanations stated above:

Artists. The 13 artists, including several who combined painting with sculpture or architecture or both, scored above the average of the three hundred eminent men in *invention*, *science*,³ *mathematics*, *handwork*, *conversation*, and *administration*.

Mathematicians. The 12 mathematicians, including 5 astronomers, scored above average in *science*, *invention*, *art*, *medicine*, *handwork*, *business*, *administration*, *philosophy*, and *public speaking*.

Musicians. The 11 musicians scored above average in *art* and in *humor*.

Non-fictional Prose Writers. The 19 non-fictional prose writers (Carlyle, Erasmus, Samuel Johnson, Voltaire, etc.) scored above average in *conversation*, *humor*, *novels*, *history*, *poetry*, *drama*, *business*, *languages*, *social theory*, *philosophy*, *politics*, *public speaking*, and *law*.

Novelists and Dramatists. Eighteen novelists (ranking second in *drama*) and 14 dramatists (ranking ninth in *novels*) scored above average in *poetry*, *music*, *humor*, *history*, *non-fictional prose*, *conversation*, *languages*, and *art*.

Philosophers. The 22 philosophers strictly so-called (ranking first in *social theory*) and 9 "social theorists" (ranking second in *philosophy*) scored above average in *mathematics*, *non-fictional prose*, *science*, *law*, *music*, *history*, *languages*, *politics*, and *administration*.

Poets. The 26 poets scored above average in *drama*, *non-fictional*

³Italics indicate a rank order of fourth or better in comparison with the other eleven types of genius, and an average score of 1.0 or better including zeros. Abilities have been listed in rank order. The artists *rank* higher in *invention* than in *science*, but it is hardly as reliable, because of the large number of zeros (no information) and consequent low average score.

prose, music, *languages*, *conversation*, *humor*, *history*, novels, art, handwork, politics, and philosophy.

Religious Leaders. The 22 religious leaders included 8 who were classed as primarily preachers, 12 who were classed as primarily writers or philosophers, and 2 who were classed as administrators. The philosophers and administrators ranked second in *public speaking*; the preachers and administrators ranked first in *philosophy*; the preachers and philosophers ranked first in *administration*. The group as a whole scored above average also in warfare, *poetry*, *politics*, music, *languages*, and invention.

Scholars. The 18 "scholars" (including 13 historians, who ranked first in *languages*, 5 philologists, who ranked fourth in *history*, and 1 lawyer) scored above average also in *administration*, *politics*, law, *philosophy*, *science*, *non-fictional prose*, drama, and music.

Scientists. The 26 scientists scored above average in *medicine*, handwork, *mathematics*, invention, *public speaking* (i.e., teaching, in most cases), *administration*, and art.

Soldiers. The 32 "soldiers" (including 18 generals, 4 admirals, 3 explorers, and 7 soldier-statesmen such as Cromwell and Napoleon) scored above average in *politics*, business, medicine, mathematics (artillery, navigation), and art.

Statesmen. The 57 statesmen scored above average in *public speaking*, law, warfare, business, *social theory*, *conversation*, non-fictional prose, humor, handwork, and history.

In general, these relationships correspond to what would be expected on the basis of common sense. Some of the more noteworthy exceptions are the high rank of artists in science and mathematics; the lack of interrelationship between mathematics and music, in either direction; the high rank of scholars, philosophers, and prose writers in politics; the high rank of statesmen, conversely, in scholarly pursuits; and the rather low rank of scientists in other scholastic pursuits.

In attempting to generalize from the preceding facts, it can be observed, first of all, that there are certain abilities which tend to appear together. For example, science, mathematics, medicine, invention, and handwork form a rather closely-knit cluster which is common to the scientists, mathematicians, and, to some extent, to the artists and philosophers. The relation of art to the science-cluster is especially interesting. Leonardo da Vinci is a supreme example, and Dürer, Michelangelo, Raphael, Rubens, Copernicus, Galileo,

Huygens, Newton, and Fulton, lesser ones, of combined artistic and scientific ability. Science, like art, usually involves both manual and visual abilities; to this extent, at least, their fields overlap.

Philosophy, social theory, history, and languages form a less closely knit cluster, allied to science on one side and to non-fictional prose on the other.

Politics, warfare, and business seem to form another cluster. Politics and warfare, at least, have had in the past a very close inter-relationship; and the statesmen and soldiers rank first and third, respectively, in business. Statesmen, however, tend to be much more scholarly than soldiers.

Finally, it is clear that novels, poetry, and drama form a compact cluster. Though it is true that the dramatists ranked low in novel-writing, the novelists ranked second in drama, the poets ranked first in drama and second in novels, and every one of the 32 novelists and dramatists was given a positive score in poetry. This cluster is rather closely allied to non-fictional prose, and all four groups are characterized by high scores in conversation, and languages, combined with low scores in science and mathematics.

Religious leadership seems to be allied with politics and administration on one side, and with scholarship on the other.

Musicians appear to have no special relationship with any other group except, possibly, the artists. Five of the 11 musicians showed artistic ability; on the other hand, only one of the 13 artists (da Vinci) was recorded as possessing musical ability.

In general, then, it can be said that there are two well-defined clusters, the *scientific* (science, mathematics, medicine, handwork, invention, and possibly art) and the *literary* (novels, drama, poetry, and perhaps non-fictional prose). There are also two less well-defined clusters, the *scholastic* (philosophy, social theory, history, languages) and the *administrative* (politics, warfare, business). The musicians are the only group that could not be made to fit into the scheme in any way.

Though he fully expected to do so, the writer found no evidence of what could be called an aesthetic cluster. The poet seems more like the novelist or essayist than like the musician or artist; the artist is more like the scientist than like the poet or musician; the musician is like no one except himself.

In addition to these minor clusters, there is one major cluster which is very much in evidence, and which may have theoretical signifi-

cance, namely, the cluster of *verbal abilities*. If a man achieved eminence in any one of several different occupations involving the use of words, he was very likely to show ability in several other activities also involving the use of words. The literary and scholastic types seem to have a great deal in common. Poet and philosopher alike were likely to distinguish themselves in the study of French or ancient history; while scientist, musician, and soldier, though one would expect them to be unlike in almost every other respect, were similar in their relative inability to manipulate words. To illustrate this tendency, let us select five types of genius which seem to involve the use of words (non-fictional prose writers, poets, novelists and dramatists, philosophers, and scholars) and eleven abilities which also seem to involve words (non-fictional prose, poetry, novels, drama, philosophy, social theory, law, public speaking, history, languages, and conversation). These types rank as shown in Table 6 in regard to the number of these abilities in which they are above the average of the whole group.

TABLE 6

Type	Number of verbal abilities		Av. corrected IQ (1, p. 84)
<i>Verbal</i>			
Prose writers	9	Philosophers	180
Poets	7	Scientists	175
Scholars	6	"Writers EHCS"	170
Novelists, Dramatists	5.5	Relig. Leaders	170
Philosophers	5	Poets, Nov., Dram.	165
		Statesmen	165
<i>Non-verbal</i>			
Statesmen	4	Artists	160
Religious Leaders	3	Musicians	160
Mathematicians	2	Soldiers	140
Artists	1		
Scientists	1		
Musicians	0		
Soldiers	0		

The correspondence is complete. Every one of the verbal types surpassed every one of the non-verbal types in number of verbal abilities shown outside its own field. Cox's IQ ranking of her types, classified somewhat differently, is given for the sake of comparison. In broad outline, the rankings correspond. In both of them the most clearly non-verbal types, artists, musicians, and soldiers, are near the bottom (cf. the ranking for "general versatility," page 465). The one

outstanding exception is the scientists, who rank near the bottom in verbal abilities and near the top both in IQ and in "general versatility."

This correspondence was to be expected, since the same information which led to the ranking of verbal abilities was used in estimating the IQ. At the same time, it suggests in a rather striking way the existence of a more or less general mental capacity, manifesting itself here in a verbal or linguistic form, and related in some way to intelligence.

DISCUSSION OF RESULTS

The first purpose of this study was "to estimate the versatility of three hundred eminent men, as an indication of the extent to which specialization is favorable or unfavorable to the attainment of eminence." If bare figures told the whole story, the answer would be decisive. We could say, not only that these geniuses were not one-sided freaks, overdeveloped on one side of their natures and atrophied on all the rest, but that they were actually far more versatile than the average college graduate of today. They were judged superior to the average graduate in 2015 instances, and inferior in only 141. Even if 30% of the positive scores were disregarded because they represent abilities which contributed to eminence, and 40% more were disregarded because they represent activities which took up only a very small amount of time (these percentages are very unreliable), there would still remain 605 positive scores in contrast to the total of 141 negative scores. Positive scores would still be more than 80% of the total (746), and negative scores less than 20%.

There is one very important possibility, however, preventing complete confidence in the verdict. The halo factor is an uncontrolled variable which quite possibly accounts for much of the difference found. It seems probable that many biographers, writing largely because of interest in and admiration for the individual studied, quite innocently suppress unpleasant facts in their efforts to present an admirable figure. The 2015 positive scores, usually based on concrete evidence, are probably relatively reliable, but the total of 141 negative scores is possibly only a fraction of what it should be. Since there are 4450 cases about which there is no information at hand, it is possible that a large number of these should be added to the negative total.

The writer himself believes that this is only a remote possibility.

The fact that the negative scores tend to be smaller and less reliable than the positive scores, the fact that positive scores still predominated in the additional information he obtained during a more intensive study of three typical cases, and the fact that each one of three men supplementing the data from their own rather extensive knowledge of biography inserted positive scores far more often than negative, all seem to show that the conclusion drawn is not fundamentally wrong. In spite of the halo factor, and the other very definite limitations of the method used—limitations which would be prohibitive if individuals and not groups were being considered—the writer believes that the excess of positive scores is about as well established as it can be by any biographical method. Crucial evidence, of course, can be furnished only by a direct study of living individuals, by men experienced in the analysis of human traits, and by tests standardized in relation to definite norms. Such evidence would probably change what is now a very high probability into almost a certainty. But its great value would come, not in verifying our main conclusion, but in testing out the specific, qualitative hypotheses discussed below.

The first of these specific hypotheses is that the versatility of genius is more a matter of ability than of interest. The ratio of positive to negative scores was 6.6 to 1 in items representing interest alone, and 21.5 to 1 in items representing definite evidence about ability or both interest and ability. This suggests that possibly, in relation to their own abilities and opportunities, the geniuses are actually narrower, more concentrated and focused, as it were, than the average man. The same theory is supported by the extremely tentative findings on pages 469-470. It appears that comparatively few of the scores represent a large amount of time. The typical genius seems to have superabundant energy combined with ease and rapidity in a broad range of activities. With such a combination, he can show creative achievement in several fields without ever endangering the one field—often a broad one—on which his fame primarily rests.

This point is a fundamental one for the theory and practice of education. Superficially considered, the versatility of genius would seem to indicate that versatility should be rather indiscriminately encouraged in gifted children; but if it is really accounted for by ability, energy, and opportunity rather than by diversity of interest *per se*, the educational implications are very different. In that case

the duty of the teacher would be, rather, to see to it that the potential genius did not fritter away his time in too great a variety of miscellaneous interests.

To determine this point, it is absolutely essential to study living individuals in order to obtain accurate records of time. In the last analysis it comes down to a question of days and hours. Does the man who achieves great things do so partially because he concentrates his days and hours on those things more than the man of mediocre accomplishment would? There is now evidence that even with such concentration he might appear to be a more versatile person, simply because the hours which he does spend on other activities are so much more productive. But the more fundamental question remains to be answered.

What kinds of ability do our data indicate, and what kinds of deficiency? The list of abilities given in Table 4 gives some indication that the "average" genius is likely to possess a kind of all-round verbal ability. The first seven items on the list (although this is an extremely rough criterion) all depend greatly on the use of words. Only one of the last seven items depends on words to the same extent. If there is any deficiency at all, it is in such relatively non-verbal activities as mathematics, music, business, and handwork. There are certainly no consistent indications of deficiency in such general categories as "social ability" (note the high rank of conversation), "practical ability" (note the 80 positive scores in politics), or "aesthetic ability" (note the high rank of poetry).

It should be remembered, however, that the "verbal type" of genius predominates in our group. What has been said above about an all-round verbal capacity probably does not apply to artists, musicians, or soldiers as much as to novelists and philosophers.

What kind of genius is the most versatile, and what kind the least? We cannot say. Our data furnish no basis for comparing the value of versatility along such diverse lines as music, conversation, politics, and mathematics; and until that is done the use of such a term as "general versatility" will be a hindrance to clear thinking, rather than an aid. We can say only that, on the basis of a crude mathematical average, our eleven musicians were decidedly less versatile than most of the other men studied; and possibly the same could be said of the artists and soldiers.

The low rank of musicians may be due to a lack of overlapping between the abilities involved in music and the abilities involved in

most other sorts of achievement. The musician may have less in common with the poet than the poet has with the philosopher. Or it may be that music is a harder taskmaster, and demands a more complete consecration, than other types of achievement. If so, there is a real danger that some of our child-musicians today are being dwarfed, musically, by the "broadening" process of public school education.

For almost all the other varieties of genius, however, the fact of versatile ability seems well established. It can be said with some assurance that they are able to carry on a surprisingly wide range of activities without seriously impairing achievement in the major field. To what extent this is a by-product of their extraordinary intelligence, and to what extent it actually subtracts from the hours devoted to their primary occupations, cannot be determined without a direct study of living individuals.

The second purpose of the study was "to discover what kinds of special ability are associated with certain kinds of genius, as an indication of the vocational types to be kept in mind in the education and guidance of gifted children." The findings are given in detail on pages 373-374, and do not require any general discussion. The existence of "ability-clusters," on the other hand, has some theoretical interest. It would, of course, be going entirely too far to claim that the scientific, literary, administrative, and scholastic clusters observed in our data correspond to general personality types. It is not claimed that they classify any aspects of personality except those entering into the choice of a vocation; but, as vocational types, they probably represent a simple and convenient classification that can be used tentatively in future work on the nature and causes of genius. The distinction between verbal and non-verbal types is especially important. If it could be verified with a large group of living eminent men, it would have a direct bearing on the much-discussed problem of "the nature of *g*."

On all of these more special questions, the need of such confirmatory evidence is painfully obvious. The reasons for it will be discussed in detail in the following section.

CRITICISM OF METHOD

Several criticisms of the method used in this study may be made, of which the most outstanding will be stated and discussed.

1. *The data are second-hand, and therefore fundamentally un-*

reliable. The writer believes this to be the most valid of the criticisms to be discussed. Biographical data of the usual type are by their very nature relatively unreliable. It is exasperating, for one who tries to hold before himself a high standard of scientific precision, to be eternally conscious that his rock-bottom facts are not scientifically precise. On the contrary, they are often snap judgments, made by untrained observers, often by prejudiced observers, without reference to established norms. Even with the most careful discrimination of good and bad sources, and with the most impartial treatment of the data so obtained, the fundamental handicap cannot be overcome to any great extent.

On this account, the writer believes that a thorough study of living eminent men, with modern testing techniques, would be far more valuable than the biographical approach here used. It would, of course, be far more time-consuming also, but it would lay a firm foundation for itself that is entirely lacking in the present study.

At the same time, one fact should be kept in mind: errors that are due to chance alone, and not to some persistent bias of observation or interpretation, can be expected to cancel each other out to a large extent when the number of separate items is more than two thousand, as it is here. The broad conclusions are far more likely to be correct than any one score given to any one man. It is in our more specific conclusions, such as the finding that our eleven musicians were less versatile than the other geniuses studied, that the danger of chance errors becomes really great.

2. *Educational and cultural conditions of the present day are so different from those of past ages that no true comparison is possible.* This is another reason, almost equally important, for verifying the tentative conclusions of the present study by a really thorough study of living eminent men. What was true in eighteenth-century France or sixteenth-century Italy may easily be false in twentieth-century America.

When this has been done, however, a comparison with the present results will be of interest not only to the psychologist, but also to the historian and sociologist. Significant changes, during the past century or two, may be found. For example, it may appear that the percentage of eminent men who write poetry (at least 40% of our 300 did so) has declined during the last century, in both Europe and the United States.

3. *The results are ambiguous because they do not differentiate*

between versatile interests and versatile ability. This difficulty is inherent in the nature of the data, and is yet another reason for verifying the results with a group in which the two factors can be more adequately distinguished. It is not so much that intense interest cannot be taken as an indication of some ability. When we read, for instance, Byron's statement that "from the moment I could read, my grand passion was history," and his voluminous list of the histories and biographies he had read before the age of 19, it would be absurd not to give him a positive score in history. But even if such statements were disregarded, the fundamental ambiguity would remain. There would still remain thousands of positive or negative scores which might have been given if the subject had shown enough interest in a given activity to demonstrate his fitness or unfitness for it. Perhaps Byron was also above average in mathematical ability, or business ability, or musical ability. We do not know, because his interest in those fields was not great enough to figure in our records.

Here again it is the more specific conclusions that are the least reliable. We are almost as sure that the typical genius surpasses the typical college graduate in range of interests as that he surpasses him in range of ability. It will be remembered that the ratio of positive to negative was found to be 7 to 1 in interest scores and 22 to 1 in ability scores. On the other hand, it is impossible to tell how far our "ability-clusters" are in reality interest-clusters, or how far the seeming non-versatility of our musicians was due simply to a lack of interest in other fields.

* * * *

Since various other criticisms have been anticipated on pages 463, 466, 469, and 472, they need not be discussed here.

1. Three hundred eminent men were judged to possess special abilities superior to those of the average college graduate in 2015 instances, and inferior in 141. Insofar as this result is not caused by the halo factor (biographers failing to record unfavorable facts), it probably indicates that the abilities of the typical genius are decidedly more versatile than those of the average college graduate of today.

2. There are some indications that the typical genius is more superior in range of ability than in range of interests.

3. The musicians, and possibly also the artists and soldiers, were less versatile than most of the other types.

4. Four "ability-clusters" (or perhaps "interest-clusters") were suggested. Of these the scientific and literary clusters are well marked; the scholastic and administrative are less well marked. There was almost no evidence of an "aesthetic type."

5. A tendency to intercorrelation of many different verbal abilities was apparent.

6. A study of living eminent men, using standard tests of some sort, is essential for real proof of the suggestions made by this investigation.

APPENDIX: CASE NOTES

To give a more definite idea of the nature of the data, condensed case notes are added for seven subjects. The first two, Goethe and Franklin, have the highest total number of points of any in the group, and may be said to indicate roughly the upper limit of human versatility. The next two, Rembrandt and Ney, are at the opposite extreme of the distribution, and represent roughly the lower limit of versatility as found in the men studied. The last three are the cases selected from the center of the distribution for further, more intensive study. They may be considered rather typical of the group as a whole.

Most Versatile Cases

Goethe

Main field of eminence: poetry. Positive scores: drama 5, novels 5, philosophy 4.5, conversation 4.5, science 4, languages 4, non-fictional prose 4, administration 3.5, art 3.5, history 3, medicine 3, handwork 3, politics 2.5, humor 2.5, law 2, social theory 2, music 1, invention 0.5. Positive scores: 18. Negative scores: none. Total points: 58.5.⁴

In poetry, novels, and drama, Goethe ranks with the greatest writers of all time. The philosophy incorporated in many of his works, though he never formulated a "system," is all-embracing in its scope. As a statesman, he was the guiding spirit of the little duchy of Weimar for more than fifty years. As a scientist, "in his work on the metamorphosis of plants and on animal morphology, he

⁴This measure has not been used at all in the body of the article.

foreshadowed the work of Darwin as no other of his contemporaries," also writing works on geology and on the psychology of color. As a friend he was charming, and had many love affairs. As an artist, he made caricatures, etched, and wrote much criticism of art and architecture. In languages, we find that at the age of 11 he was learning Hebrew on his own initiative, and that at the age of 16 he was writing verses in French, English, and Italian. Not to mention his ability in medicine, handwork, social theory, music, or invention, we may add that "in the political and legal history of Germany his knowledge extended to minute details."

Franklin

Main field of eminence: politics (including diplomacy). Positive scores: science 5, non-fictional prose 5, humor 5, conversation 5, business 4.5, administration 4.5, philosophy 4, social theory 4, invention 4, handwork 3, poetry 2, public speaking 1, drama 1. Positive scores: 13. Negative: none. Total points: 48.

A statesman and a diplomat of the first order, Franklin was also an eminent scientist at a time when science in America hardly existed, and a delightful writer at a time when "the stilted, verbose and turbid habit was tediously prevalent." His educational experiments embodied theories that were generations ahead of his time. He was also an expert printer, a very able financier, a master of propaganda, and "perhaps the most agreeable conversationist of his age."

(Although there are no other names that could easily challenge Goethe's right to first place, there are several that could challenge Franklin's to second. Jefferson, da Vinci, Galileo, Hugo, Voltaire, Constant, Beaumarchais, and Alexander Hamilton are next to him in total points, and a different importance attached to individual items might easily give any one of them the second place.)

Least Versatile Cases

Rembrandt

Main field of eminence: art (painting and etching). Positive scores: none. Negative scores: Languages —1.5, business —2.5. Total points: —4.

Though 6 works are listed in the bibliography, and there are 16 typewritten pages in the manuscript biography, the data are really relatively scanty. Little is known of Rembrandt's private life, and

almost nothing of his childhood. Perhaps he would appear more versatile if more were known.

It was said that he "proved but an indifferent scholar" at school, and that he "seems to have had little taste for reading, to judge by the small number of books to be found in the inventory of his effects in later life." Since scholarship and reading in Holland imply linguistic ability primarily, he was given a negative score in languages. The low score in business was given because he was said to be "a child in his relations with the world outside his doors," and because we have definite evidence that he was declared bankrupt at the age of 50 and spent the remainder of his life in poverty. At the same time, it should be noted that he was pre-eminent both in etching and in painting, and that this is not taken account-of in our scoring.

Ney

Main field of eminence: warfare. Positive scores: none. Negative scores: law —1.5. Total points: —1.5.

In Ney's case the lack of data is much more apparent than in that of Rembrandt; only four works are listed in the bibliography, and Cox's manuscript biography consists of only seven pages. Ney's education was only rudimentary. He then tried to take up law, but gave it up in disgust, as nothing but warfare could satisfy his craving for adventure. Eventually he became one of Napoleon's generals.

(Rembrandt and Ney are the only men whose total of points was negative. Three others—the painter Van Dyke and the musicians Palestrina and Mozart—maintained an even balance of positive and negative points. The other 295 were all positive.)

Average Cases

Davy

Main field of eminence: science (chemistry). Positive scores, before intensive study: invention 3, poetry 3, philosophy 3, non-fictional prose 2, novels 2, public speaking 2, art 1.5, history 1.5, languages 1, medicine 1. Negative scores: conversation —1.5. Number positive: 10. Number negative: 1. Total points: 18.5.

Sir Humphrey Davy was an English scientist, famous for his discovery of "laughing gas," for his work in establishing the analogous nature of chlorine, fluorine, iodine, and for his invention of the miner's safety lamp. His career as a poet began at the age of five, and he wrote verses throughout his life. Coleridge, who was an

intimate friend, once said that if Davy had not been the first chemist, he would have been the first poet of his age. The score of 3 in philosophy is based on a thick notebook which he kept at the ages of 16 to 19, crammed full of metaphysical discussions. He was scored 2 in non-fictional prose because he was said to have "a force of eloquence (in his scientific writings) which could issue only from a mind of the highest powers and of the finest sensibilities." He was scored 2 in novels because of the wonderful and terrible tales he used to invent, as a boy of 8. His "extraordinary popularity as a lecturer" accounts for the score of 2 in public speaking; his rather crude paintings of birds, fishes, and landscapes account for the score 1.5 in art; his extensive reading in history accounts for the score of 1.5; the fact that he learned to speak French, apparently in not much more than a year, accounts for his score of 1 in languages; and his work in medicinal chemistry accounts for the score of 1 in medicine. But, since it was said that he had a "brusqueness and superciliousness due to an ungraceful timidity which he could never conquer," he was scored —2 in conversation. This was questioned by Dr. Terman, and as a result the score of —2 was changed to —1.5.

Three of these scores were changed, and three added, as a result of the more intensive follow-up. The score of 1 in languages was changed to —1 because of the discovery that he never did learn to speak French fluently or pronounce it correctly, though he lived for a short time in France. The score in non-fictional prose was raised from 2 to 3 by the discovery of two books of semi-philosophical essays which he wrote in later life. The score of —1.5 in conversation was raised to 1 by the discovery of a great deal of material tending to show that he was a vivacious, and at times an eloquent, conversationalist. A new score of 1 was given in mathematics because he taught himself the fundamentals of the subject at the age of 18; a new score of 1 in administration because for seven years he was president of the Royal Society, ultimately unpopular, but fairly efficient; and a new score of —4 in music, because it was once said that his friends could not even teach him the air of "God Save the King." The net result of this revision was to change his point score from 18.5 to 18.0.

Spinoza

Main field of eminence: philosophy. Positive scores: social theory 4, science 3.5, handwork 3, languages 3, mathematics 2.5, conver-

sation 1.5, medicine 1. Number positive: 7. Negative: none. Total points: 18.5.

Spinoza's writings on free speech and on the theory of government rank with his best work. A tolerable physicist, and a pioneer in biblical criticism, his scientific temper is shown also in the subject-matter and the mathematical structure of his philosophy. He obtained his living by grinding lenses, and was a "proficient optician." He knew Spanish, Portuguese, Hebrew, Latin, and some Greek, as well as enough Dutch for simple conversation with his fellow-countrymen; possibly also French, German, and Italian. Although extremely quiet and frugal, taking almost no time for recreation, he had several intimate friends, and "people of culture felt a peculiar charm in his presence."

As a result of the intensive follow-up three new scores were added, but none of the original ones were changed. A score of 1 in law was given because of his knowledge of Hebrew law; a score of 1 in non-fictional prose was given because "his library was as rich in *belles lettres* as it was poor in philosophy;" and a score of 3 in art because he was said to be "an accomplished draughtsman, and left at his death a portfolio full of sketches which he had drawn for his own pleasure." The net result of the revision was to increase the total of points from 18.5 to 23.5.

Webster

Main field of eminence: politics. Positive scores: public speaking 5, law 4, conversation 3.5, social theory 3, history 2, languages 2, poetry 2, novels 1. Negative scores: handwork —1, business —1.5. Number positive: 8. Negative: 2. Total points: 20.

Daniel Webster's reputation as a lawyer rests upon his defense of the principle of nationalism in the Dartmouth College case, etc. He was said to be "greatly distinguished for his conversational powers and genial temper in society." The score of 2 in languages rests upon his translation of two law volumes from Latin and Norman French; the score of 2 in poetry, upon the fact that some of his companions (age 17) thought he should be a poet; the score of 2 in history upon his very extensive reading in this field; and the score of 1 in novels on "moderately extensive reading in English generally." But, because of his admission, "somehow I could never learn to hang a scythe," he was scored —1 in handwork; and be-

cause he "habitually lived beyond his means," he was scored —1.5 in business.

The follow-up resulted in conflicting statements about his manual ability, which may be considered to cancel each other and leave the previous score of —1 unchanged. The previous score of 3.5 in conversation was reduced to 2.5 in the light of his college roommate's statement that he was "not very popular with the class," and the lack of confirmatory evidence for the statement on which the original score was mainly based. A new score of 1 in science was given in view of the statement that "minute observation of nature" was one of his strongest characteristics. The net result of the revision was to leave the previous total of 20 points unchanged.

REFERENCE

1. Cox, C. M. *Genetic studies of genius: Vol. 2. The early mental traits of three hundred geniuses.* Stanford University, Calif.: Stanford Univ. Press, 1926. Pp. 842.

*Stanford University
California*

LA DIVERSITÉ DES APTITUDES CHEZ LES HOMMES DE GÉNIE

(Résumé)

Pour estimer la valeur de la spécialisation, et aussi pour contribuer à une classification inductive du génie en termes de "groupes d'aptitudes," on a fait une étude des aptitudes de 300 hommes éminents dans les domaines autres que leurs domaines spéciaux. On a obtenu les données dans les biographies en manuscrit rédigées par C. M. Cox, lesquelles ont été plus complètes de beaucoup que les données qu'elle a publiées. Ces faits ont été ensuite évalués par deux personnes. Les corrélations entre elles-mêmes de leurs évaluations en rapport aux traits spécifiques ont donné un moyenne de $0,76 \pm 0,02$. L'évidence obtenue indique que le génie est décidément d'esprit souple. On a jugé que les 300 hommes possèdent des aptitudes spéciales supérieures à celles du diplômé universitaire moyen américain en 2015 cas, et inférieures en 141 cas. On a noté des aptitudes différentes telles que musique, politique, histoire, drame, mathématiques, et conversation. Dans la plupart des cas l'évidence a suggéré l'intérêt aussi bien que l'aptitude. On a tiré la conséquence que l'aptitude d'un homme éminent est ordinairement d'une nature si générale qu'il peut cultiver des intérêts en plusieurs domaines sans nuire à son oeuvre dans son propre domaine. On a trouvé aussi que quatre "groupes d'aptitudes" semblent exister. Parmi ceux-ci les groupes scientifique et littéraire sont bien marqués; les groupes "scolaire" et administrateur sont moins marqués. On n'a qu'une petite tendance à la corrélation entre elles-mêmes des aptitudes "esthétiques", mais il s'est montré une corrélation entre elles-mêmes de toutes les aptitudes verbales.

WHITE

DIE VIELSEITIGKEIT DES GENIES

(Referat)

In der Absicht, den Wert der Spezialisierung abzuschätzen und auch zu einer induktiven Klassierung des Genies als Gruppe von Anlagen (ability clusters) betrachtet beizutragen, untersuchte man die Fähigkeiten von 300 hervorragenden Männern ausserhalb des Hauptfeldes ihrer Leistungen. Die Daten wurden in den in Manuscript-Form bestehenden, durch C. M. Cox gesammelten Biographien gefunden, welche viel ausgiebiger waren, als die von ihr publizierten Tatsachen. Jene Daten wurden dann von zwei Abrechnern erwertet. Die durchschnittliche Interkorrelation zwischen ihren Erwertungen mit Bezug auf spezifische Züge (traits) war $.76 \pm .02$. Die ermittelten Daten deuten an, dass Genie entschieden vielseitig ist. In 2015 Fällen hielt man, dass die 300 Männer Sonderfähigkeiten (special abilities) besaßen, in Bezug auf denen sie dem durchschnittlichen Graduierten einer Amerikanischen Universität überlegen waren, und in 141 Fällen wurden sie in Bezug auf diese Sonderfähigkeiten als unterdurchschnittlich betrachtet. So verschiedene Fähigkeiten wie sie Musik, Politik, Geschichte, Drama, Mathematik, und Sprachverkehr (conversation) darstellen wurden eingetragen. In den meisten Fällen schloss die Fähigkeit auch Interesse in sich ein. Man folgert, dass die Fähigkeiten eines hervorragenden Menschen gewöhnlich so allgemeiner Natur sind, dass er Interessen verschiedener Artew nachgehen kann, ohne seinen Leistungen in seinem Hauptfelde zu schaden. Man fand auch, dass es vier Gruppen von Anlagen zu geben scheint. Von diesen sind die wissenschaftliche und die literarische Gruppe scharf ausgeprägt. Die "scholastische" und die verwalterische (administrative) sind es weniger. Die Korrelationen unter den "ästhetischen" Fähigkeiten erwiesen sich als unbedeutend, aber alle sprachliche (verbal) Fähigkeiten erwiesen sich als unter einander bestimmt korreliert.

WHITE

SHORT ARTICLES AND NOTES

THE INFLUENCE OF INTELLIGENCE AND PERSONALITY TRAITS UPON FALSE BELIEFS

FRANCIS F. POWERS

The purpose of this article is to describe the technique and results of an inquiry into the relations obtaining between intelligence, tendency to false beliefs (credulity), and certain personality traits, more particularly, introversion.

The whole subject of false beliefs and superstitions is a very interesting, not to say practical, one. That superstitions vary considerably in their factual basis is quite apparent. That they further influence everyday behavior in varying degrees is also obvious and it is from this fact that superstitions derive their importance in education. Whether intelligence is a significantly determining factor in the extent to which our behavior patterns are influenced by false beliefs is not so patent. Further, the relative susceptibility of various personality types of subscription to superstitions is not established.

The experiment herewith described details the results of correlating measures of intelligence, introversion, and superstition. A few preliminary assumptions and reservations may be made:

First, the difficulty of devising a good test of false belief or superstitions. It is recognized that many present-day superstitions originally had and to some extent still retain a utilitarian value. An effort was made to include only such false beliefs as bore a remote and not obvious connection with reasonable actions.

Secondly, the difficulty of securing a candid response upon a superstition test. College students seem to desire the aspect of sophistication. By a technique later described, rather satisfactory results were secured.

Thirdly, the recognized absence of high correlation between various measures of introversion. This has been demonstrated by Guthrie (1). The low correlations found, however, demonstrate either that the term introversion is misleading in its implication of some unit personality trait, or that, if such a trait exists, it can be measured only by a test of rather restricted scope. For the purposes of this investigation, extroversion is defined in terms of degree of correspondence to commonality of response in an association test. The test employed was the one devised by Whately Smith and described in "The Measurement of Emotion" (2).

With the foregoing limitations in mind, the procedure here described was followed:

A test of 75 of the most prevalent superstitions was compiled. The items

were taken partially from Dresslar's *Superstition and Education*, and the list was arranged in order of popularity—the most common listed first. An attempt was made to include only superstitions and exclude, as far as possible, misconceptions and false beliefs which might have some basis in present fact or scientific belief. They were worded in such a way as to get affirmative answers if the individual's behavior was influenced by a slight belief or fear as well as if there were an acknowledged complete belief. This was not an easy task, and it is possible that some claimed non-belief who are in reality influenced by either a belief that they had at some other time or a caution because of belief held by associates. The items were presented in the form of questions and the subjects were asked to underline either "Yes" or "No" in answer. In case of non-belief the answer underlined would be in every case "No."

A test for specific traits of personality was not so easy to form, and it was finally decided to use a free association test for traits of introversion and extroversion which had been arranged by Whately Smith. In this test are listed one hundred common words which are presented at intervals of ten seconds and the association response noted. Commonality of response is taken to denote traits of extroversion. When the response given by the subject was the same as the most commonly offered response it was checked. The score obtained in this way is called the "score of extroversion."

Since the tests were to be used on undergraduate students at the University of Washington and since an intelligence test is given to each freshman student who enters the University, these scores were available to use as intelligence scores. The test used was the University of Washington Intelligence Entrance Examination, devised by Dr. William R. Wilson. Its reliability is .78 (1, p. 87).

Three first-term psychology classes, comprising, in all, 97 individuals, were used as subjects for the test. Sixty-eight of the group were freshmen and most of the others sophomores. The tests were presented toward the end of May, near the close of their first year of college and their first term of psychology. Although no count was kept of the sexes, there were slightly more women than men. The tests were presented to all groups the same day, at nine, one, and two o'clock, respectively.

About 20 minutes were allowed to check the test on superstitions, but the time limit was used only to prevent any comparison of papers. The test on superstition was given first with no instructions except those appearing at the top of the page, a further admonition to be candid, and a statement that the tests had no bearing on the class standings or grades of the subjects. The words in the association test were pronounced, one each ten seconds, and the subjects wrote down their responses. They were asked to submit their first response, and it is fairly likely that they did since so little time was allowed.

In scoring the papers in the test for superstitious belief, the "yes" answers were counted and serve as a raw score.

Finally, correlations were computed between:

1. Intelligence and superstition scores
2. Intelligence and extroversion scores
3. Superstition and extroversion scores

The reliabilities of the tests used were as follows:

1. Wilson's Intelligence Test .79 (1000 cases)
2. Superstition Test .83 (97 cases)
3. Extroversion Test .80 (210 cases)

The above reliabilities represent self-correlations (odd scores correlated with even scores) corrected and raised by the application of Brown's formula.

The correlation between intelligence and extroversion is .067. This substantiates the conclusion of Guthrie who found low intercorrelation between various measures of introversion-extroversion and intelligence.

The correlation between intelligence and superstition is $.122 \pm .08$. This figure happens to correspond to the highest correlation which Guthrie obtained between any two tests of extroversion (1).

The correlation between superstition and introversion-extroversion is $.103 \pm .06$. [Range on introversion 3-39 (100 items); superstition 0-36 (76 items).] When correlated for attenuation the foregoing correlations become:

1. Intelligence and extroversion .09
2. Intelligence and superstition .15
3. Superstition and extroversion .12

Within the limitations of the technique the foregoing results would seem to indicate:

1. That there does not exist a high degree of relationship between any two of the factors being measured. Previous work has demonstrated that the so-called introverted type of mind is neither caused by, nor a concomitant of, any particular level of intelligence. The results here found further indicate that intelligence would not appear to be a safeguard against credulity or false belief.

2. That the introvert, insofar as he can be measured at present, is no more or no less inclined to superstition than the extrovert. None of the correlations, simple or corrected, is sufficiently high to contain the slightest predictive value. (Note: The coefficients of alienation are all approximately .99, and, since a correlation of .866 is necessary before the error of estimate has been reduced a half, the results here found contain no element of accurate prediction.)

Both introversion and tendency to conduct based on false beliefs seem to be results of early training and learning rather than to be related either to each other or intelligence.

REFERENCES

1. GUTHRIE, E. R. Measuring extroversion and introversion. *J. Abn. & Soc. Psychol.*, 1927, 22, 82-88.
2. SMITH, W. The measurement of emotion. London: Kegan Paul; New York: Harcourt, Brace, 1923. Pp. 184.

*University of Washington
Seattle, Washington*

THE RELATIONSHIP BETWEEN KNOWLEDGE AND A MEASURE
OF AUTISTIC THINKING ON CERTAIN INTERNATIONAL
PROBLEMS

WILLIAM W. BIDDLE

Autistic thinking has been variously defined as uncritical, unreal, emotional, prejudiced, or wishful thinking. It has been one of those generalized concepts, coming from an older psychology, which we have had difficulty in making specific. Autistic thinking, as it occurs in pathological cases, is easily discernible. But, as we move over from the so-called abnormal, to the normal, it is much more difficult to determine which thinking is realistic and which is not. We are often forced into the position of condemning as autistic those conclusions which do not agree with ours. Undoubtedly, there is much thinking among normal people which is autistic. However, if the term is to have any value for an objective psychology, it must be objectively discernible and, if possible, measurable.

Autistic thinking is essentially wishful in nature. It is the lonesome child who creates an imaginary playmate. It is the neurotic with definite inferiorities who has delusions of grandeur. By saying that the thinking is wishful, we mean that its conclusions are satisfying to the organism. The unreal conclusion is accepted, because it fits in comfortably with certain preconceptions. We more readily believe that which is emotionally satisfying, that which we wish to believe.

In attempting to objectify and measure the autistic tendency in thinking, it is possible to make use of this willingness to accept conclusions which are in accord with previous biases and emotional thinking. Goodwin Watson's Test of Fair-Mindedness was a beginning in this field.

It is probable that a great majority of the widely accepted thinking on topics of general social interest is autistic. Particularly in the realms of social conflict, religion, industrial problems, international relations, and the like, thinking is more than likely to be unreal and uncritical. Often, the social scientist, on entering these fields, finds himself, willy-nilly, a party to the controversy, drawn up in a battle array against opposing camps. There is great need for objective measures of the autistic tendency in the field of social thinking.

A test for measuring this tendency in certain international relations of the United States has been devised. It consists of 14 short articles, quotations from, or adaptations of, a variety of controversial material bearing on the problems. The articles, taken from books, periodicals, speeches of congressmen, and propaganda leaflets, run the gamut of discussion from obvious, vituperative prejudice to reasonable, objective weighing of the issues. The subject is asked to rate each article on a graphic rating scale as to its probable "fairness" and "truth" (even though he may be unfamiliar with certain of the facts quoted). Judges have been used to establish a criterion rating as a scoring device. The tendency toward autistic thinking on this problem is indicated by the amount which any subject may deviate from the criterion. The extent to which a subject deviates may be regarded as an index of his willingness to accept conclusions which fit in comfortably with biased preconceptions. A total autistic score is obtained by summing and averaging all deviations.

The purpose of a test of this nature has to be disguised. In this case, the test is justified as an ordinary opinion ballot. In order to give credence to this statement, the subject is also asked whether or not he agrees with the opinions expressed in each article. In addition, he is asked to record, in a sentence or two, the impression which the article gives him. This recording of impressions will be noted when we discuss validity.

The type of test described is more cumbersome than the ordinary true-false or multiple-choice type. It had to be limited, therefore, not only to the field of international relations, but also to certain specific problems within that field. For reasons extraneous to this paper, the test confines itself to the international problems of the United States centering about the Pacific Ocean. All conclusions reached bear upon the Pacific problems only. It may prove valuable to investigate tendencies toward autistic thinking as it concerns other areas of international conflict by this technique. As far as this test is concerned, however, it appears to the subject as an opinion ballot on certain international relations of the United States which center about the Pacific Ocean. The whole study must be regarded as a preliminary investigation of the relationship between knowledge and prejudiced reactions within a restricted area of international discussion.

The reliability of the test as described is .872, correlating the odd items against the even and using the Spearman-Brown formula (N equals 100). When we come to discuss validity, however, we find that it is difficult to make a statistical statement. That subjects are giving a response to emotionally toned, biased, or autistic material is empirically observable. The extent to which subjects may have "seen through" the test and given responses other than they would ordinarily give is partially indicated by the reaction to the question, "How does this article impress you," etc. Out of 400 cases, only two subjects indicated by their written reactions that they were suspicious of the purposes of the test. Theoretical validity would be .933 (the square

root of the reliability). It should be noted, of course, that this test does not pretend to be a measure of autistic thinking in general. It is a test of autistic tendencies on certain international problems in a literary situation.

The measure of knowledge of international relations which was used in this study is of the conventional true-false, multiple-choice type. It is concerned with knowledge about the international relations of the United States, in general, not merely the problems of the Pacific.

The correlation between scores on the test of knowledge and scores indicating autistic thinking is $-.357 \pm .043$. That is, as knowledge on general international problems increases, there is a small tendency for autistic thinking, at least on the problems of the Pacific, to decrease. Conversely, there is a small tendency for critical thinking to increase as knowledge increases. Approximately 200 subjects were used. They are public-high-school seniors in New York City; Norwalk, Connecticut; Detroit, Michigan; and college freshmen in Lehigh University, Bethlehem, Pennsylvania. All correlations are the Pearson r , uncorrected for attenuation.

When we come to examine the autistic reactions within individual areas in the Pacific situation, we find some interesting comparisons. Correlations between knowledge and autistic reactions within three areas are as follows:

The Philippine question	$r = -.027 \pm .048$
The Chinese question	$r = -.207 \pm .046$
The Japanese question	$r = -.266 \pm .068$

The correlation between knowledge and reactions toward the Philippine situation is practically zero. Does this indicate that these subjects are more emotional in their thinking about the Philippines and less affected by information in an area in which they are strongly emotional? Further investigation would be needed to answer this question.

The correlations between knowledge and reactions within the Chinese and Japanese areas are sufficiently similar to warrant comparisons. The mean autistic score for the Japanese situation is 28.32 and for the Chinese, 30.9. The difference, 1.58, is not statistically significant. However, it is interesting to note that the autistic scores grow out of a sympathy for China and an antagonism for Japan. This becomes clear if we mark deviations from the criterion favoring the Oriental nation as plus, and deviations that disapprove that nation as minus. Fifty tests, chosen at random and scored in this manner, indicate a mean score of $+27.3$ for Chinese items and -50.3 for Japanese. The difference, 77.6, which is 3.88 times the standard error of the difference, may be taken as characteristic of the whole group. Why autistic reactions of these subjects should be favorable toward the Chinese and strongly antagonistic toward the Japanese is a question which may be answered in terms of recent political events. It is curious, in view of this difference, that correlations of autistic reactions with knowledge should be as close as .207 and .266.

This technique of measuring autistic reactions may be adaptable to other material within the field of international relations. It also might be worthy of consideration as a measure of prejudice on industrial, religious, or racial questions.

In view of the preliminary and specific nature of this study, it would be difficult to generalize results. However, if further correlations between knowledge and other types and areas of autistic thinking should prove as low as ours, it would be obvious that other factors than the intellectual are involved in the achievement of critical or realistic thinking. Perhaps we shall have to educate the emotions as well as the intellect.

*Western Reserve University
Cleveland, Ohio*

ARE GIRLS SUPERIOR TO BOYS IN VISUAL MEMORY?

PAUL L. BOYNTON

This investigation was made as a result of a position taken by Meumann in his *Psychology of Learning* (Baird's translation), where that author followed studies such as those by Pohlmann, Cohn, and Wreschner to conclude that girls are superior to boys in visual memory. From a biological standpoint it was not evident why this superiority of one sex over the other should exist because, to exist for biological reasons, of course, there would necessarily have to be a sex-linkage of some type, and, while this might be possible, it hardly seemed probable. Pohlmann, himself, as noted by Meumann, did not assume that this superiority of the girls in visual memory, to which Meumann referred, was inherent to the sex but rather that it was the product of training. It was decided that some light might be thrown on the problem if several different types of materials were used in an investigation; the results of such a procedure are reported herein.

Seven different types of tests were used with a total group of 90 girls and 135 boys, students in the Department of Psychology, at the University of Kentucky. The first test consisted of 20 colored pictures, taken primarily from advertisements, which appeared in certain well-known magazines for women. The pictures were pasted on 8½- by 11-inch cardboards and were exposed to view at the rate of one every five seconds. Immediately after the completion of the showing of these 20 pictures, a second group of pictures, taken from the same sources, were shown, these at the rate of one every ten seconds. In this latter group of 20 were the exact duplicates of 10 pictures which were shown in the first 20. As each picture of the latter group was shown, the students recorded whether or not they thought they had seen it in the first 20 pictures. These pictures were taken from women's magazines because it was thought that they would likely have more appeal

to women than to men. In other words, a definite attempt was made to get pictures which should be of more interest to girls than to boys.

The second test was comparable to the first except that here the 18 pictures which compose the Du Pont game-bird series were used, and, instead of showing two whole series, only 9 pictures were shown at first, and then the entire 18 were shown later, with the students having to record after each of the 18 pictures was shown whether or not they thought it was shown in the first group of 9. This series was chosen with the feeling that the theme of the series, game birds, would likely have more interest for boys than for girls, but it is quite true, as anyone familiar with these pictures will recognize, that there are certain aesthetic and nature-study features connected with the pictures which makes it difficult to say whether the total interest factor would be greater for boys than for girls.

The remaining five tests were presented to the students through the use of a lantern and trans-lux screen. The first of these, actually the third test, was a test with 20 nonsense syllables, divided into two parts of 10 syllables, each. After the first set of 10 was shown, the students wrote all the syllables they could remember, and the same procedure was duplicated with the second set of 10.

The fourth test was divided into two parts. Part A consisted of the names of 20 kinds of dress-materials, clothes, or adornments used by girls. Part B was of the same nature except the materials all pertained to boys. In the fifth test, 5 sets of *X*'s and *O*'s were shown to the students and then 10 other sets were shown. In this latter group of 10 were to be found the first 5 sets together with 5 other sets. The students were to identify the 5 sets they saw in the beginning. These sets were of such nature as the following: *OXXO*, *XOXXO*, *XXOOXOOX*. The sixth test consisted of 10 sets of four-digit numbers, such as 7526, which were first exposed on the screen and then replaced by a second set of 10, 5 of which were also found in the first series. The seventh test consisted first of a picture containing 14 geometric figures. These figures were of unusual curvilinear design, diverging very perceptibly from the type of figures used in ordinary geometric work. After the presentation of these first 14, a second picture was shown, containing 14 more figures, 7 of which were also in the first picture. The students were to identify those which they saw in the first picture of 14.

As is quite evident from a review of the tests, no one method of presentation of the materials to be learned was adhered to throughout. This was purposely avoided in order not to make method the factor of much importance. The methods used for the first, sixth, and seventh tests were similar in that a series of material was presented and half of this material was also presented for identification in a second series. There is one difference of importance, though, and that is that the sixth and seventh tests were presented on the screen whereas the first was not. Also, the

length of each of these tests, or the amount of material to be learned, was different for each of the tests. The second and fifth tests were comparable, but here, also, one was presented on the screen and one was not, and one had two series of 9 and 18 parts, whereas the other had two series of 5 and 10 parts. The third and fourth tests probably were more nearly alike as regards method of presentation than were the others. Recognition was not a factor in the response on either of these tests, since no stimulus words or cues were used in either case. Where certain similarities in methods of presentation did exist it was thought that the radically different types of materials used would offset this effect.

It is quite evident from a casual glance at Table 1 that girls made a better general record on the tests than did boys. On only two of the eight tests were the boys' medians as high as the girls'. It should be noted, however, that in only two cases, Test I and IV-A, are the differences as much as four times their probable errors, and in both of these tests the materials were definitely chosen because of their supposed appeal to girls. In other words, the only two cases of unequivocal superiority of the girls over the boys are at points where an effort was made to handicap the boys and aid the girls.

It cannot be said that boys took advantage of their opportunities in the

TABLE 1

MEDIANS FOR BOYS AND GIRLS, AND MEDIAN DIFFERENCES BETWEEN BOYS AND GIRLS, TOGETHER WITH THE ERRORS OF EACH MEASURE, FOUND ON EACH OF THE SEVEN TESTS OF VISUAL MEMORY

Test	Boys		Girls		Girls	Differences		
	Median	<i>P.E.</i> _{md.}	Median	<i>P.E.</i> _{md.}		<i>P.E.</i>	Boys	<i>P.E.</i>
I	16.95	.14	18.27	.18	1.32	.23		
II	15.87	.21	15.85	.21			.02	.30
III	12.92	.23	14.11	.29	1.19	.37		
IV-A	16.08	.17	18.12	.13	2.04	.21		
IV-B	17.21	.21	18.12	.15	.91	.26		
V	8.48	.14	8.42	.16			.06	.21
VI	7.42	.11	7.85	.14	.43	.18		
VII	11.14	.12	11.48	.16	.34	.20		

same manner as did girls. On IV-B, which was the test of memory for things pertaining to the boy, girls did better than did boys, the difference being .91, with a probable error of .26. Upon further consideration it seems wholly possible that this may simply reflect a greater interest on the part of girls in the general "clothes situation," and this would appear to be testified to by the girls' medians, which are 18.12 in both situations. Whereas the girls' medians would indicate that the two tests are of equal difficulty, the boys' medians indicate that the B part of the test is signifi-

cantly easier than the A part. The difference between the boys' medians on the two parts is 1.13, with a probable error of the difference of .27. Here, again, then, would seem to be evidence that the type of material used is of fundamental importance in determining the existence or extent of sex differences in visual memory.

The difference of .02 favorable to the boys on Test II is actually an indication of similarity rather than difference. While it was supposed that boys would have more interest in the theme of these game-bird pictures than girls, as previously mentioned, there are factors connected with the pictures other than simply that of sporting activities. Color harmonies, color clashes, and other aesthetic elements, as well as the factor of nature study, might well counterbalance the advantage which the boys were supposed to have on this test. In any event, it is important to note that when the materials were changed from those which so obviously favor girls to those where girls do not enjoy so obvious an advantage, sex differences disappear.

Test III, nonsense syllables, yields results favorable to girls. While the median difference, slightly over 3 times its probable error, is not sufficiently large to be absolutely unequivocal in meaning, it does point with fair probability to the superiority of girls. No definite reason can be assigned for this difference, but it seems possible that it may indicate a greater patience on the part of girls with the "nonsense" situation.

The differences existing between boys and girls are so small on Tests VI and VII that one would be unwise to attempt to conclude that they are indicative of any definite relationships. If taken in conjunction with Tests I, III, IV-A, and IV-B, it is true that these two tests may add to the cumulative effect in such a manner as to be of some significance, but individually the differences on both tests are too small, when compared with their respective errors, to have any outstanding meaning.

When Test V was prepared it was devised with the idea that it was more nearly a new, or novel-situation test than any other one in the group. Also, it was thought that for both sexes there would be fewer pre-formed attitudes toward this *XO* material than toward that used in the other tests. With this basic assumption in mind, it is interesting to note that sex similarity on this test is much more pronounced than is sex difference. A difference of .06 with a probable error of .21, can have only one interpretation and that is, "lack of difference." It is especially important that this "no difference" conclusion shall have appeared on the test which was devised for the purpose of presenting material where previous training or habits were thought not to be important factors in the determination of the response.

More girls go to beauty parlors than do boys. More boys go to baseball games than do girls. These sex differences exist without question. At the same time, however, they are easily explicable in terms of the specific

training and the general environmental influence to which the two sexes are exposed. None would be so radical as to suppose that the differences have a *direct* chromosomal cause. Similarly, there are probably sex differences in visual memory, and in many fields of measurement girls are probably superior to boys. Such differences as do appear, however, are apparently not due to any inherent difference in the sexes, but rather to differences in the experiences of the sexes, as expressed through the materials used in measurement. This study would seem to indicate that, while it is possible to show that girls are superior to boys in visual memory, it is also possible to show that girls are not superior to boys in visual memory. Therefore, when specific differences are obtained they should be examined in light of the most pertinent explanatory factors and not be taken as bases for broad generalizations.

George Peabody College for Teachers
Nashville, Tennessee

BOOKS

JOHN E. ANDERSON, CHARLOTTE BÜHLER, ANNA FREUD, ARNOLD GESELL, FLORENCE L. GOODENOUGH, LETA S. HOLLINGWORTH, SUSAN ISAACS, HAROLD ELLIS JONES, MARY COVER JONES, VERNON JONES, C. W. KIMMINS, HEINRICH KLÜVER, KURT LEWIN, HELEN MARSHALL, DOROTHEA MCCARTHY, MARGARET MEAD, JOSEPH PETERSON, JEAN PIAGET, RUDOLF PINTNER, LEWIS M. TERMAN, BETH L. WELLMAN, AND HELEN T. WOOLLEY. [CARL MURCHISON, ED.] *A Handbook of Child Psychology*. Worcester, Mass.: Clark Univ. Press; London: Oxford Univ. Press, 1931. Pp. xiv+711.

This volume satisfies its title, for it is a manual which contains twenty-two chapters dealing with twenty-two different topics pertaining to child psychology, and written by the same number of authors. The topics themselves cover a wide range: from the extremely descriptive and practical chapter on "Eating, Sleeping, and Elimination," by Helen T. Woolley, to one on "The Eidetic Child," by Heinrich Klüver, which, by contrast, is theoretical and evaluative; from "The Gifted Child," by L. M. Terman, on the one hand, to "Psychoanalysis of the Child," by Anna Freud, and "Children's Dreams," by C. W. Kimmins, on the other. Each chapter is followed by an extensive bibliography of materials indicated in the text and of materials not so used.

The papers are, for the most part, surveys of what the authors believe to be representative or outstanding studies, included under categories into which they have divided their respective subjects. Naturally, some topics are more nearly complete and more competently handled than others. Learning in Children," by Joseph Peterson, in which a relatively small number of studies are analyzed in some detail, while they are also evaluated and interpreted, is, I believe, among the best chapters of the volume. Gesell's chapter on "The Developmental Psychology of Twins," including a large number of studies in the survey, is valuable for its comprehensiveness. Unfortunately, however, not all the papers possess the merit of Peterson's or Gesell's (and others); some are very sketchy, while others are simply condensed repetitions of what the authors have themselves published elsewhere. Examples of chapters open to the latter criticism are Goodenough's on "Children's Drawings" and Isaacs' on "The Experimental Construction of an Environment Optimal for Mental Growth"; while "Children's Plays, Games, and Amusements," by Helen Marshall, is open to the former criticism.

Not all of the authors have limited themselves to surveys of the literature or to description only. Kurt Lewin, for example, presents a theoretically interpretative chapter dealing with the influence of environmental forces on development and their pedagogical values in training. Though we may question the values of his adoption of certain terms of physics (such as vectors, valences, etc.) and their application to his problems, there is little

doubt of the penetration and significance of his approach. Another instance is "The Primitive Child," by Margaret Mead which, while not very illuminating—due, no doubt, to the paucity of materials in the field—is interesting and significant for its suggestions of problems and methods of research in social psychology, which might yield information not otherwise obtainable.

On the whole, there seem to be three general types of studies in the volume: those which are entirely survey, or very nearly so; those which combine the survey with interpretation and evaluation; and those which are principally theoretical in nature. The frequencies fall in that order, which seems reasonable in a handbook. Important, also, is the fact that the student of child psychology has available here a very valuable fund of source materials, in addition to the essentials which have been extracted by some of the authors and the contributions of others. The volume is, perhaps, the most comprehensive one in the broad and challenging field of child psychology.

Cornell University

FRANK S. FREEMAN

ACKERSON, L. Children's behavior problems: a statistical study based upon 5000 children examined consecutively at the Illinois Institute for Juvenile Research: I. Incidence, genetic and intellectual factors. (*Behav. Res. Fund Monog.*) Chicago: Univ. Chicago Press, 1931. \$4.00.

CARLSON, H. S. Information and certainty in political opinions: a study of university students during a campaign. *Univ. Iowa Stud.: Stud. Char.*, 1931, 4, No. 1 (New ser., No. 209). Pp. 48. \$1.00.

FRYER, D. The measurement of interests in relation to human adjustment. New York: Holt, 1931. Pp. xxxvi+488. \$3.60.

KAHN, E. Psychopathic personalities. New Haven, Conn.: Yale Univ. Press, 1931. Pp. xii+521. \$5.00.

KANNER, L. Judging emotions from facial expressions. *Psychol. Monog.*, 1931, 41, No. 186. Pp. iv+194.

MURCHISON, C. [ED.] Anderson, J. E., Bühler, C., Freud, A., Gesell, A., Goodenough, F., Hollingworth, L. S., Isaacs, S., Jones, H. E., Jones, M. C., Jones, V., Kimmins, C. W., Klüver, H., Lewin, K., Marshall, H., McCarthy, D., Mead, M., Peterson, J., Piaget, J., Pintner, R., Terman, L. M., Wellman, B. L., and Woolley, H. T. *A handbook of child psychology*. Worcester, Mass.: Clark Univ. Press; London: Oxford Univ. Press, 1931. Pp. xiv+711. \$5.00.

VOLD, G. B. Prediction methods and parole: a study of factors involved in the violation or non-violation of parole in a group of Minnesota adult males. Hanover, N. H., Minneapolis, Minn., Liverpool, Eng.: Sociological Press, 1931. Pp. xiv+138.

SUBJECT INDEX

- Achievement** and forgetting of the pleasant and unpleasant, 216
- Addiction** to drugs and birth order, 52
- Affective experiences** and forgetting, 216
- Alcoholism** and birth order, 52
- Animal incentives**, 263
- Audience as incentive**, 277
- Autistic thinking** *vs.* knowledge in international problems, 493
- Belief** and intelligence and personality, 490
- Biologist**, psychological reactions of, 102
- Birth order** and behavior, 52
- Change** in social attitude, measurement of, 230
- Character measurements**, definitions in, 114
 - rating and family size, 22
- Chinese**, performance and Binet intelligence of, 402
- Color preferences** of Japanese children, 297
- Cooperativeness** and family size, 10
- Cultural background** and behavior, 54
 - and family size, 9
- Czechoslovak psychology** of testing, 238
- Defect** and inferiority attitude, 87
- Definitions** in character measurement, 114
- Delinquency** and birth order, 52
- Depression** and perseveration, 46
- Distraction** as incentive, 280
- Emotion** of actor and empathy, 235
- Empathy** and actor's emotion, 235
- Enuresis** and birth order, 52
- Explosive personality** and birth order, 52
- Family size** and personality, 3
- Forgetting** of the pleasant and unpleasant, 216
- Genius**, versatility of, 460
- Honesty** and family size, 18
- Human incentives**, 269
- Idiots**, problem-solving among, 306
- Implements**, use of by idiots, 306
- Incentives**, animal, 263
 - human, 269
 - psychology of, 261
 - in religious experience, 392
- Indifference function**, 139
- Inferiority attitude** and organic inferiority, 87
- Inhibition** and family size, 17
- Intelligence** of adults, measurements, 72
 - of Chinese on performance and Binet tests, 402
 - and false belief, 490
 - and forgetting of the pleasant and unpleasant, 216
 - linguistic *vs.* non-linguistic ability, 245
 - of parents and family size, 21
 - tests, requirements of, 72
 - and versatility, 460
- Interests**, sex differences, 120
- Introversion** and perseveration, 46
- Japanese**, color preferences of, 397
- Knowledge** *vs.* autistic thinking on international problems, 493
 - of results as incentive, 269
- Linguistic** *vs.* non-linguistic ability, 245
- Measurement** of adult intelligence, 72
- Measures** of perseveration, 30
- Memory**, early, quantitative study, 252
 - and intelligence and achievement, 216
 - racial comparisons, 205
 - sex differences in immediate and delayed, 433
- Moral knowledge** and family size, 8
- Motion pictures**, sex differences in recall of, 433

- influence on children's social attitude, 230, 291
- Motivation in religious experience, 392
- Motor inhibition and perseveration, 35
- Music as incentive, 281
- Neuroticism and early memories, 252
- Opinions of college students, 168
- Organic inferiority and the inferiority attitude, 87
- Performance *vs.* Binet tests with Chinese, 402
- Perseveration as a general factor, 28
- and introversion, 46
- Persistence and family size, 16
- Personality and birth order, 52
- and false beliefs, 490
- and family size, 3
- and handwriting, 337
- Praise and reproof as incentives, 273
- Problem-solving among idiots, 306
- Punishment as incentive, 275
- Racial comparisons in recall, 205
- differences in color preference, 397
- preferences of university students, 179
- Recall, racial comparisons, 205
- sex differences in immediate and delayed, 433
- Religious bodies, census, notes on, 417
- experience and temperament, 374
- opinions of college students, 168
- Reward as incentive, 275
- Rivalry as incentive, 279
- Satisfaction function, 139
- Service and family size, 13
- Sex differences in conversational interests, 120
- in color preference, 400
- Social attitude, measurement of change in, 230, 291
- behavior and handwriting, 337
- conformity, note on attitude of, 126
- Socio-economic opinions of college students, 168
- Temperament and religious experience, 374
- Testing in Czechoslovakia, 238
- Tests of intelligence for adults, 72
- Versatility of genius, 460
- Withdrawn personality and birth order, 52
- Writing movements and social behavior, 337
- Zero-point in intelligence tests, 74

